PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE LETRAS

CRISTIANE RUZICKI CORSETTI

CONVERSATIONAL COMPETENCE IN ENGLISH AS A SECOND LANGUAGE: A STUDY OF PRAGMATIC MARKERS

DOCTORAL THESIS

PORTO ALEGRE

2015

CRISTIANE RUZICKI CORSETTI

CONVERSATIONAL COMPETENCE IN ENGLISH AS A SECOND LANGUAGE: A STUDY OF PRAGMATIC MARKERS

Doctoral thesis submitted as a prerequisite to obtain the PhD degree from the Post-graduation programme of Faculdade de Letras da Pontifícia Universidade Católica do Rio Grande do Sul.

Advisor: Professor Cristina Lopes Perna
Supervisor: Professor Jonathan Culpeper

PORTO ALEGRE

2015

CRISTIANE RUZICKI CORSETTI

CONVERSATIONAL COMPETENCE IN ENGLISH AS A SECOND LANGUAGE: A STUDY OF PRAGMATIC MARKERS

Doctoral thesis submitted as a prerequisite to obtain the PhD degree from the Post-graduation programme of Faculdade de Letras da Pontifícia Universidade Católica do Rio Grande do Sul.

Accepted on 8 January, 2015.

EXAMINATION BOARD:

_____

Dr. Cristina Lopes Perna - President - PUCRS

_____

Dr. Jonathan Culpeper - Lancaster University

_____

Dr. Marília dos Santos Lima - Unisinos

_____

Dr. Simone Sarmento - UFRGS

_____

Dr. Karina Veronica Molsing - PUCRS

I dedicate this thesis to my parents, my husband and my children, to whom I am grateful for their endless support and encouragement.

ACKNOWLEDGEMENTS

I would like to thank the following people and institutions, without whom I would not have been able to accomplish my academic objectives:

Professor Jonathan Culpeper, for his invaluable guidance, counselling, thought-provoking discussions and generosity.

Professor Cristina Lopes- Perna, for her constant support and counselling.

Professor Jorge Campos da Costa, for his supervision at earlier stages.

Dr. Dana Gablasova and Dr. Carmen Dayrell at CASS (Centre for Corpus Approaches to Social Sciences), Lancaster University, for their advice on corpus investigations.

CAPES Foundation, Ministry of Education of Brazil, for sponsoring my PhD studies at PUCRS and at Lancaster University.

FALE Post-graduation Department at PUCRS, for their support.

The subjects of this study, who kindly agreed to participate in the empirical project.

My son André, for his careful revision of the formatting of this thesis and his unconditional support.

José, Mariana and Rodrigo, for their understanding and support.

RESUMO

A conversa é uma das atividades mais básicas da comunicação verbal. Indivíduos transmitem informações, realizam ações, estabelecem e mantêm relações sociais, entre outros atos, ao engajarem-se em trocas conversacionais. A conversa geralmente não é planejada, ocorre em tempo real e envolve reciprocidade. Esta tese apresenta os fundamentos da Análise do Discurso, da Pragmática e da Análise da Conversa, detalhando fenômenos discursivos, pragmáticos e interacionais que entrelaçam-se durante conversas presenciais. A seguir, foca no aprendiz de línguas, abordando a Pragmática da Interlíngua e discutindo modelos de competência comunicativa e definições de competência conversacional. Esta tese justifica e propõe um modelo de competência conversational em segunda língua, no escopo da Pragmática, que constitui-se de três componentes: o gerenciamento do discurso, a negociação do significado ilocucionário e a implementação de práticas conversacionais. Este estudo inclui investigações, baseadas na Linguística de Corpus, de marcadores pragmáticos característicos dos componentes propostos. Visando os objetivos desta tese, criou-se um pequeno corpus especializado, com a produção oral de aprendizes brasileiros no nível CEFR B1. Subcorpora orais oriundos do "The BNC Sampler" e "The Diachronic Corpus of Present-Day Spoken English" foram empregados como corpora de referência de inglês britânico. Os capítulos empíricos desta tese analisaram os advérbios mais comuns utilizados para mediar segmentos de discurso em conversas, os "hedges" adverbiais explícitos e implícitos mais comuns utilizados para mitigar atos de fala representativos e as partículas de resposta mínimas utilizadas pelo interlocutor para expressar uma boa receptividade. As investigações enfocaram os marcadores pragmáticos "*well*", "*really*", "*actually*", "*maybe*", "*probably*", "*just*'', "*yeah*" e "*uhuh*". Concluiu-se que os sujeitos deste estudo se beneficiariam de práticas pedagógicas visando a aquisição de marcadores de discurso em geral, das funções pragmáticas do ajustador "*just*" e de formas mais variadas para expressar uma boa receptividade. Os marcadores "*really*" e "*yeah*" foram empregados, em suas funções discursivas e pragmáticas, com frequências adequadas pelos aprendizes brasileiros. O advérbio "*maybe*" foi sobre-utilizado, sinalizando uma tendência para a utilização de formas adverbiais para expressar epistemicidade, ao invés de verbos modais.

Palavras- chave: Competência Conversacional. Marcadores Pragmáticos. Investigações baseadas em Linguística de Corpus. CEFR B1. Aprendizes Brasileiros.

ABSTRACT

Conversation is one of the most fundamental activities in verbal communication. When people engage in conversational exchanges, they transmit information, perform actions, establish and maintain social relationships, among other acts. Conversation is often unplanned, takes place in real time and involves reciprocity. This thesis presents the foundations of Discourse Analysis, Pragmatics and Conversation Analysis and details discourse, pragmatic and interactional phenomena which seem to be intertwined during face-to-face conversations. It then focuses on the second language learner by addressing Interlanguage Pragmatics and by critically revisiting communicative competence frameworks and definitions of conversational competence. It justifies and proposes a conversational competence model in L2 within a pragmatic domain, comprising three facets: the management of discourse, the negotiation of illocutionary meaning and the deployment of conversational practices. It includes corpus-based investigations of pragmatic markers which are prototypical of the proposed facets of conversational competence. A small specialised corpus of Brazilian learners´ oral production at CEFR B1 had been previously built for the purposes of this thesis. Spoken sub-corpora derived from "The BNC Sampler" and "The Diachronic Corpus of Present-Day Spoken English" were selected as British English benchmark corpora. The empirical chapters of this thesis examined the most common discourse marking adverbs used to mediate segments of discourse in conversations, the most common explicit and implicit adverbial hedges used to mitigate representative speech acts and the most common minimal response tokens used to express good listenership. The pragmatic markers chosen for investigation were "*well*", "*really*", "*actually*", "*maybe*", "*probably*", "*just*'', "*yeah*" and "*uhuh*". It was concluded that the subjects of this study would benefit from pedagogical assistance for the acquisition of discourse markers in general, for the pragmatic functions of the adjuster "*just*" and for more varied forms to express good listenership. The markers "*really*" and "*yeah*" were employed, in their discourse and pragmatic functions, with adequate relative frequencies by the Brazilian learners. The adverb "*maybe*" was overused, signalling a tendency to use adverbial forms to express epistemic stance instead of modal verbs.

Key-words: Conversational Competence. Pragmatic Markers. Corpus-based Investigations. CEFR B1. Brazilian Learners.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

INTRODUCTION

Conversation is one of the most fundamental activities in verbal communication. When people engage in conversational exchanges, they transmit information, they perform actions, they establish and maintain social relationships, among other acts. When speakers express propositions, they are usually aware that some portions of their knowledge are truer or more accurate than others and that not all knowledge is equally reliable. Thus, they qualify such knowledge with expressions which indicate their assessment of its degree of reliability, the likelihood of its being a fact (Chafe, 1986). They also mark their stance or attitude to the message or to preceding discourse. In addition to expressing propositions, speakers use language to get other people to do things, to express psychological states and to change the world around them (Searle, 1976). In conversations, speakers display their public self-image and employ politeness strategies in order to minimise the degree of imposition of acts contrary to hearers' wants and to convey common ground (Brown & Levinson, 1987).

Conversation is often unplanned, takes place in real time and involves reciprocity. By conversation, I refer to "small talk, casual conversations and more formal conversations between two or more speakers which take place in real time in face-to-face situations". Speakers are expected to keep talking, they have little time to plan ahead and they modify or readjust their discourse (Biber et al. 1999). In other words, speakers produce spoken discourse, monitoring the understanding of intended meanings, and adjust it accordingly. Making meaning is a dynamic process, involving the negotiation of meaning between the speaker and the hearer, the context of an utterance and its meaning potential (Thomas, 1995). Hearers play their role, signalling understanding (or lack of), convergence and engagement and encouraging speakers to continue their ongoing turns.

Within the Brazilian educational context, English and the theme "conversational competence" have never been more in vogue. With the "Science without Borders Programme", Brazilian undergraduate students are given the opportunity to study at an overseas university for up to a year, as part of their graduation scheme. In order to be eligible for funding, applicants need to sit a formal language examination. Among the most popular examinations requested by institutions are the multi-level exams "International English Language Systems" (IELTS) and "Test of English as a Foreign Language" (TOEFL). Some institutions also accept certificate

exams such as “Cambridge English First” (FCE) and “Cambridge English Advanced” (CAE), which are level-specific. As for results, candidates will usually require a successful performance at levels B2 or C1 of the “Common European Framework of Reference for Languages” (CEFR), depending on the demands of the target country institution and on the examination taken per se.

The CEFR (Council of Europe, 2001) provides a common basis for syllabus and examinations design, curriculum guidelines and textbooks. It details what abilities language students have to learn and to develop in order to be able to use language for communicative purposes and to act effectively. It defines levels of proficiency which enable learners’ progress to be measured at each stage of the learning process and on a life-long basis. The CEFR is based on six broad reference levels and is an action-oriented approach to language teaching and learning. This approach views language users and learners as social agents who perform tasks in specific circumstances, environments and within particular fields of actions. Both CEFR B1 and B2 levels are within the independent user category and, in a broad sense, correspond to the intermediate and upper intermediate levels, respectively. CEFR C1 and C2 levels represent the proficient user category.

At the very beginning of the “Science without Borders Programme” in 2011, a considerable number of Brazilian students chose to apply to Portuguese universities, due to the difficulties they encountered in communicating in English or in other languages. As a consequence, the Brazilian government created another programme called “English without Borders” in 2012. This programme offers undergraduates at federal universities the opportunity to take a levelling test, to engage in distance learning activities and to attend English classes so as to improve their overall linguistic ability.

Generally speaking, Brazilian learners of English tend to find the listening and speaking components of international examinations particularly challenging. My main motivation comes from my teaching experience preparing students for IELTS and FCE, targeting CEFR B2. Learners at CEFR B1 usually have difficulties understanding pragmatic meaning as, many times, they are not able to recognise contextual clues that lead to the making of inferences, since they are still struggling with coded information. Corsetti’s findings (2009) showed that her learners found the interpretation of the illocutionary force of indirect speech

acts particularly difficult, as forms and functions do not necessarily correspond. In addition, some learners at CEFR B1 are unable to move beyond the intermediate plateau in terms of spoken production.

In my Master's dissertation (Corsetti, 2009), I critically analysed two communicative competence constructs (Canale & Swain, 1980, Canale, 1983 and Bachman, 1990), of which pragmatic competence is a sub-category. In both frameworks, spoken discourse and written discourse are not dealt with separately. However, oral and written discourse differ tremendously vis-à-vis the demands they make on language producers, their forms and their composite features. Thus, the ability to produce spoken discourse and interact with other speakers is complex enough to justify a separate characterisation from the ability to produce written discourse, especially if we consider the negotiation of meaning which takes place in real time through spoken discourse. In addition, second language learners perform language functions (Bachman & Palmer, 1996) and convey and negotiate meanings through spoken discourse, relying on the complex interplay between discourse, pragmatic and interactional features.

The aims of this thesis are three-fold:

1. To revisit the construct "conversational competence" in L2 and place it within a pragmatic domain, grounded on the notion of "meaning in interaction" (Thomas, 1995);
2. To detail key discourse, pragmatic and interactional phenomena which comprise conversational competence in L2;
3. To investigate some of the key discourse, pragmatic and interactional phenomena comprising conversational competence in the oral production of twenty Brazilian learners of English at CEFR B1.

As conversational competence is a broad concept encompassing a number of different phenomena, I have chosen to analyse the functions of pragmatic markers that lend themselves to corpus investigation and seem to be prototypical of the three proposed facets of conversational competence (despite some possible overlap), namely discourse marking adverbs, adverbial hedges and backchannels. While pragmatic markers have little or no

propositional meaning, they inform us of the pragmatic relations between speakers, messages and contexts (Culpeper & Kytö, 2010). Discourse markers are a subtype of pragmatic markers which have the additional feature of mediating between one speaker's utterance and another (Culpeper & Kytö, 2010). Hedges are modifying devices that speakers use to mitigate or soften the force of their messages (Nikula, 1996). Backchannels are vocal indications used by listening partners to indicate that they are following what is being said and also to inform how the message is being received (Yngve, 1970).

Based on empirical data, I address four research questions, which are pursued through more specific sub-questions:

RQ1: *How do Brazilian learners at CEFR B1 mediate segments of discourse in conversations?*
RSQ1: *What are the most common discourse marking adverbs used to mediate segments of discourse in conversations by Brazilian learners at CEFR B1? How do their frequency of use and pragmatic functions compare to those of native speakers?*

RQ2: *How do Brazilian learners at CEFR B1 mitigate representative speech acts?*
RSQ2: *What are the most common explicit and implicit adverbial hedges used to mitigate representative speech acts by Brazilian learners at CEFR B1? How do their frequency of use and pragmatic functions compare to those of native speakers?*

RQ3: *How do Brazilian learners at CEFR B1 express good listenership?*
RSQ3: *What are the most common minimal response tokens used to express good listenership by Brazilian learners at CEFR B1? How do their frequency of use and pragmatic functions compare to those of native speakers?*

RQ4: *What are the pragmatic markers which seem to pose a challenge to Brazilian learners of English at CEFR B1?*
RSQ4: *Based on the findings of research sub-questions 1 to 3, what are the pragmatic markers which seem to pose a challenge to Brazilian learners of English at CEFR B1?*

In order to accomplish the objectives of this study and pursue its research questions, this thesis is divided into 6 chapters. In chapter 1, I describe the foundations of Discourse Analysis, Pragmatics and Conversation Analysis and detail discourse, pragmatic and interactional phenomena which seem to be intertwined during face-to-face conversations. By presenting an overview of these areas and their comprising features, I aim at preparing a theoretical basis for a conversational competence model in a second language, to be proposed in chapter 2, and for the empirical investigations in chapters 4, 5 and 6. I initially present different views on Discourse Analysis, features of spoken discourse and the organisation of spoken discourse, namely cohesive relations and discourse markers. Secondly, I address the scope of Pragmatics and summarise key pragmatic phenomena derived from Speech Act Theory (Austin, 1962, Searle, 1969, 1975, 1976), Conversational Implicatures (Grice, 1975, Levinson, 2000) and Politeness (Brown & Levinson, 1978, 1987). I also investigate the role of hedges as pragmatic markers. Thirdly, I describe the foundations of Conversation Analysis and detail interactional phenomena within turn and sequence organisation. Conversation Analysis is included in the literature review for its constructs but it is not used as a methodology.

In chapter 2, I focus on the second language learner by presenting Interlanguage Pragmatics and competence frameworks which are embedded in the interface between Second Language Acquisition (SLA) and Pragmatics. Firstly, I introduce views on the domains of Interlanguage Pragmatics and areas of interlanguage investigations. I also summarise SLA theories that account for pragmatic learning. Secondly, I focus on SLA more specifically by critically revisiting communicative competence frameworks (Canale & Swain, 1980, Canale, 1983, Bachman, 1990) and definitions of interactional competence (Schmidt, 1983) and conversational competence (Richards & Sukwiwat, 1985, Thornbury & Slade, 2006, Corsetti, 2009). Next, I justify and propose a conversational competence model in L2 within a pragmatic domain, highlighting key discourse, pragmatic and interactional phenomena previously presented. Lastly, I relate the proposed model of conversational competence in L2 to the level descriptors of the Common European Framework for Language Reference (Council of Europe, 2001) and justify theoretical and methodological choices.

In chapter 3, I detail the methodology employed for the empirical investigations. I initially present different views on the status of Corpus Linguistics and address learner corpora,

corpus-based investigations and techniques. Next, I discuss combined approaches to the investigation of spoken discourse, namely Corpus Pragmatics (Jucker, 2013) and Corpus Linguistics and Conversation Analysis (O'Keeffe and Walsh, 2012). In the final part of chapter 3, I describe the empirical study of this thesis, its subjects, instruments and corpora and the methodological choices and procedures employed for my investigations of pragmatic markers.

The subsequent empirical chapters are structured in a similar fashion and include a brief review of the literature related to the micro-elements under investigation. In chapter 4, I focus on the most common discourse marking adverbs used to mediate segments of discourse in conversations by Brazilian learners at CEFR B1. I initially present corpus-based views on discourse marking adverbs. I then list the most common adverbs in the Brazilian learner corpus which seem to operate as discourse markers and justify my choices for empirical investigations, namely "*well*", "*really*" and "*actually*". In the sub-sections, I investigate the pragmatic functions of each of the selected adverbs in native speaker and learner corpora and compare frequencies.

In chapter 5, I examine the most common explicit and implicit adverbial hedges used to mitigate representative speech acts by Brazilian learners at CEFR B1. I briefly revisit explicit and implicit pragmatic force modifiers (Nikula, 1996) and then present corpus-based evidence (O'Keeffe et al., 2007) of forms traditionally associated with hedging. Next, I introduce the possible candidates for hedges in the Brazilian learner corpus, discuss their main functions and justify choices for adverbial selections, namely the pragmatic markers "*probably*", "*maybe*" and "*just*". In the sub-sections, I investigate the pragmatic functions of the selected adverbial hedges. I present the Brazilian learner corpora data (position, frequency and functions), qualitative analyses and comparisons with the native speaker benchmark corpora (BNC Sampler spoken sub-corpus and face-to-face conversations from the 1990s sub-corpus) and the Spanish component of LINDSEI.

In chapter 6, I investigate the most common minimal response tokens used to express good listenership by Brazilian learners at CEFR B1. I initially address the distinction between minimal and non- minimal response tokens. I then present corpus-based findings vis-à-vis the distribution of forms and functions of response tokens in different varieties of English

(O'Keeffe & Adolphs, 2008). Next I introduce the possible candidates for minimal response tokens in the Brazilian learner corpus and justify choices for the selection of forms, namely "*yeah*" and "*uhuh*". In the sub-sections, I detail the contexts of usage of the items under investigation. I specify their distribution among functions as response tokens and illustrate them through qualitative analyses. I also compare relative frequencies with native speaker and Spanish learner data (LINDSEI).

All empirical chapters 4, 5 and 6 include final sections whose purpose is to highlight the pragmatic markers which seem to pose a challenge to Brazilian learners of English at CEFR B1. In these sections, I summarise the main points derived from specific investigations, discuss findings in the light of proposed research sub-questions and make some pedagogical recommendations. The findings of this study are intended to shed some light on the intermediate plateau phenomenon, providing teachers with an authentic sample of Brazilian learners' strengths and weaknesses in terms of spoken production at CEFR B1. A cautionary note to add is that some elements under investigation displayed low raw frequencies in the Brazilian learner corpus. As a consequence, my comments can only be indicative of general patterns. Nonetheless, the pedagogical recommendations proposed may be of some value for teachers whose students display similar difficulties.

Outside the EFL classroom, learners communicating in a second language may interact with individuals from their own speech community, with native speakers of the target language and with other learners of varied nationalities. Following from this, a question arises as to whose pragmatic norms are to be followed. Considering the notion of "English as an International Language", McKay (2002, p.24) claims that for a language to be considered "international", it will have developed to an extent in which it is "no longer linked to a single culture or nation but serves both global and local needs as a language of wider communication". Thus, I follow McKay's methodological position, which indicates that the teaching of English should focus on enabling learners to be flexibly competent in international communication through the medium of English, rather than being based on the native speaker model. However, considering learners' needs for a successful performance at international examinations, native speaker corpus data can serve as a benchmark for spoken production since it reflects authentic language in use.

Lastly, the proposed model of conversational competence summarises the functions of key discourse, pragmatic and interactional phenomena in conversations. Second language teachers may consider these descriptions when deciding on a syllabus for a programme focusing on speaking sub-skills. In addition, the qualitative analyses of this study are intended to demonstrate the interplay between discourse, pragmatic and interactional phenomena in the different facets of conversational competence. They also offer a pragmatic view on language in use. As Culpeper and Haugh (2014, p. 11) rightly state, "interaction is where pragmatic phenomena happen".

1. A REVIEW OF THE LITERATURE

Face-to-face conversations, whether in formal or informal speech events, are prototypical of language in use. However, "language in use" is a rather broad concept as it can be applied to different approaches to language. Discourse Analysis, Pragmatics and Conversation Analysis investigate instances of language in use focusing on different relationships and phenomena. In this chapter, I describe the foundations of each area and detail discourse, pragmatic and interactional phenomena which seem to be intertwined during face-to-face conversations. By presenting an overview of such areas and their comprising features, I aim at preparing a theoretical basis for a conversational competence model in a second language, to be proposed in chapter 2 and for subsequent empirical investigations, to be detailed in chapters 4 to 6.

In the first part of this chapter, I discuss different views on Discourse Analysis (Brown & Yule, 1983, Fairghclough, 1995, Van Dijk, 2008) and then focus on spoken discourse (Brown & Yule, 1983, Sinclair & Coulthard, 1975, Culpeper & Kytö, 2010). Next, I investigate the organisation of spoken discourse by addressing Halliday and Hassan's treatment of cohesion (1976) and the organisational and pragmatic functions of discourse markers (Halliday & Hasan, 1976, Hölker, 1991, Brinton, 1996, Fraser, 1996, Carter & McCarthy, 2006). In the second part of this chapter, I present different views on the scope of Pragmatics (Levinson, 1983, Thomas, 1995, Culpeper & Haugh, 2014) and summarise key pragmatic phenomena derived from Speech Act Theory (Austin, 1962, Searle, 1969, 1975, 1976), Conversational Implicatures (Grice, 1975, Levinson, 2000) and Politeness (Brown & Levinson, 1978, 1987). I also investigate the role of hedges (Lakoff, 1972, Chafe, 1986, Brown & Levinson, 1987, Nikula, 1996) as pragmatic markers. In the third part of this chapter, I describe the foundations of Conversation Analysis (Schegloff, 2007, Hutchby & Wooffitt, 2008) and detail interactional phenomena within turn organisantion (Sacks et al., 1974, Jefferson, 1986, Yngve, 1970) and sequence organisation (Schegloff & Sacks, 1973, Pomerantz, 1984, Schegloff, 2007).

1.1. DISCOURSE

The term "discourse" is ambiguous and it is used in social and linguistic research in a number of inter-related yet different ways, overlapping, at times, with the notion of language. Within

Linguistics, discourse is traditionally defined as "language above the sentence or above the clause" (Stubbs, 1983) or "language in use" (Brown & Yule, 1983). Baker (2006) explains that it is possible to talk about the discourse structure of particular texts since certain lexical items and grammatical structures are expected to appear at specific places within specific texts. The term discourse can also be applied to describe different types of language use or topics such as political discourse and media discourse (Fairghclough, 1995). Baker (2006) clarifies that language is one way in which discourses are constructed. However, language, both as an abstract system (phonetics, grammar and lexicon) and as a context-based system of communication, is not the same as discourse though we analyse language in texts in order to unveil traces of discourse. Furthermore, different types of discourse such as spoken discourse (Biber et al., 1999, Carter & McCarthy, 2006, Culpeper & Kytö, 2010) and learner discourse (Aijmer, 2009) have also been investigated in corpus-based studies[1].

Within the linguistic domain, Brown and Yule (1983) claim that the analysis of discourse relates to the analysis of language in use. Therefore, it must include the description of linguistic forms along with the purposes or functions those forms are designed to serve in real life situations. The authors adopt a functional approach to discourse analysis and detail two views that may overlap and do not completely exclude each other. The transactional view acknowledges that language may be used to fulfil communicative purposes but regards the communication of information as its most important function. The authors (1983, p. 2) refer to the language used to convey factual or propositional information as "primarily transactional language". Speakers or writers pursue the efficient transference of information. Thus, the language used for this function is message-oriented and it is important that the recipient get the informative detail correct. For instance, when a pedestrian gives directions to a tourist, it is important that the former make his or her message clear to the latter. Linguists, philosophers of language and psycholinguists have traditionally investigated the use of language for the transmission of factual or propositional information.

The interactional view focuses on language used to establish and maintain social relationships. The phatic use of language[2] has been addressed in sociological and

[1] Corpus-based research will be discussed in chapter 3.
[2] "Phatic" refers to communication between people which is not intended to seek or convey information but has the function of establishing or maintaining social contact (Richards, Platt, Weber, 1985: 214). It is one of Jakobson's functions of language (1960).

anthropological literature, particularly the conventional use of language to open and close talk exchanges. Conversational analysts have examined the use of language to negotiate role-relationships, peer-solidarity, the exchange of turns in a conversation and face saving acts (Brown & Levinson[3], 1978, Sacks et al., 1974). According to Brown and Yule (1983, p. 3), it is undeniable that "a great deal of everyday human interaction is characterised by the primarily interpersonal rather than the primarily transactional use of language". In fact, a great deal of casual conversations portray samples of language that appear more to be intended as contributions to a conversation, i.e. small talk, than to be considered as instances of information giving. Brown and Yule (1983) add that written language is used primarily for transactional purposes, although there are written genres whose purpose is not to inform but rather to maintain social relationships such as "thank you" letters, love letters and games of consequences. Likewise, text messages and online chats display features of spoken discourse used for both transactional and interactional purposes.

Within the sociocultural domain, Critical Discourse Analysis (Fairclough, 1995) is an analytical framework for investigating the connections between language, power and ideology. It claims that "naturalised implicit propositions of an ideological character are pervasive in discourse, contributing to the positioning of people as social agents (Fairclough, 1995, p. 23). They include aspects of ideational meaning, for instance, implicit propositions needed to infer coherent links between sentences, and assumptions about social relationships underlying interactional practices such as turn-taking systems and politeness conventions. Fairclough (1995) conceptualises power in terms of asymmetries between participants in discourse events and in terms of unequal capacity to control how texts are produced, distributed and consumed in particular sociocultural contexts. Thus, a range of text properties are considered as potentially ideological such as features of vocabulary and metaphors, grammar, presuppositions and implicatures, politeness conventions, speech-exchange systems, generic structure and style.

In a socio-cognitive approach, Van Dijk (2008, p. 116) defines discourse as "any form of language use manifested as (written) text or (spoken) talk-in –interaction, in a broad semiotic sense". This includes visual structures and gestures, facework and other semiotic signs for spoken interaction, combinations of sounds and visuals in many hybrid multimedia discourses

[3] This date refers to Brown and Levinson's first version of Politeness (1978).

(e.g. movies, television, mobile phones, the internet and other channels of communication). He proposes a multidisciplinary theory of context and details the relationships between context and language, context and cognition and context and discourse. His theory is grounded on the following principles: contexts are participant constructs or subjective definitions of interactional or communicative situations; they are unique constructs, featuring the ad hoc, embodied experiences of ongoing perceptions; they are mental blocks; they are a specific type of experience model; context models are schematic; contexts control discourse and comprehension; they are socially based, dynamic and often, largely planned (Van Dijk, 2008, pp. 15-18).

As for the relationship between context and discourse, Van Dijk (2008) argues that the main function of contexts (in the sense of context models and participant constructs) is to enable and to constrain the production and the comprehension of text and talk. Van Dijk (2008) claims that context models control discourse by controlling its possible variations. In other words, given the subjective model of an event, sociocultural knowledge, or group attitudes and ideologies, context models demonstrate how speakers formulate specific or general beliefs at all levels of discourse. Following from this, style is the context-controlled way in which discourse may vary and adapt to different situations whereas register is the way in which grammar plays a role in such situations. Thus, "register is the routinised way language users bring to bear grammatical resources to express the semantic and pragmatic dimensions of discourse" Van Dijk (2008, p. 222).

To conclude, we have seen that the term "discourse" is ambiguous and may be applied to different domains. Within linguistic perspectives, discourse can refer to "language in use" (Brown & Yule, 1983), "language above the sentence or clause" (Stubbs, 1983) or specific types of language use or topics. In a sociocultural view, Critical Discourse Analysis (Fairclough, 1995) investigates the connections between language, power and ideology. In a socio-cognitive domain, discourse refers to as "any form of language use manifested as (written) text or (spoken) talk-in –interaction, in a broad semiotic sense" (Van Dijk, 2008, p. 116).

In the next sub-section of this chapter, I narrow my discussion to spoken discourse since the primary object of investigation of this thesis is face-to-face conversations. I initially present

the demands spoken and written discourse make on language producers (Brown & Yule, 1983). I then focus on the structure of conversational exchanges (Sinclair & Coulthard, 1975). In the final part, I detail the composite features of spoken conversation (Culpeper & Kytö, 2010, Leech, 2000, Biber et al., 1999).

1.1.1. Spoken discourse

Brown and Yule (1983, pp. 4-5) compare the different demands spoken and written language make on language producers. Although communicative messages can be aided by paralinguistic features such as facial expressions, gestures, voice quality effects, the production of oral language seems to be more challenging. Speakers must monitor what they have just said and determine whether it matches their intentions. When they utter their next phrase and monitor that, they simultaneously plan their next utterance, fit that into the overall pattern of what they want to say, and also monitor both their performance and reception by hearers. Neither do speakers have permanent records of what they have said earlier nor do they have notes to remind them of what to say next. Writers, on the other hand, are able to examine what they have already written, pause between each word with no fear of being interrupted by interlocutors, take their time to choose lexical items, check their progress with notes and edit their texts. While speakers are expected to complete their turns within the allotted time, writers are characteristically under no such pressure. Once words are uttered, they are bound to be heard by interlocutors and if those words fail to convey speakers' intended messages, active public repair will have to be undertaken. Conversely, written words can easily be erased or modified before being published without causing any major consequences.

Nonetheless, Brown and Yule (1983) indicate that speakers have a few advantages over writers. Based on the observation of interlocutors' responses and/or facial expressions, speakers are able to modify what they are saying so as to make their discourse more accessible or acceptable to hearers. Writers have no immediate feedback and need to predict and guess readers' reactions. Speakers have the advantage of monitoring their interlocutors' immediate reaction but suffer from the disadvantage of exposing their feelings and of having to speak clearly and concisely. They also need to immediately respond to whichever way interlocutors react.

Now that we have compared the different demands spoken and written language make on language producers, let us focus on spoken discourse in particular. In a sociolinguistic investigation, Sinclair and Coulthard (1975) attempt to unveil the structure of discourse. They initially choose to analyse classroom conversational exchanges between teachers and pupils in situations where the teacher was at the front of the class "teaching" and thus, possibly exerting the maximum amount of control over the structure of discourse. Sinclair and Coulthard (1975) account for the structure of conversational exchanges based on a taxonomy of classes of conversational moves. According to Sinclair and Coulthard (1975, p. 44) "moves are made up of acts, and moves themselves occupy places in the structure of exchanges". Following from this, the structure of moves is described class by class. There are five classes of moves (framing, focusing, opening, answering and follow-up) which realise two classes of exchanges (boundary and teaching). "Framing" and "focusing" moves realise "boundary" exchanges whereas "opening", "answering" and "follow-up" moves realise "teaching" exchanges.

With regards to the functions of these moves, "framing moves" are used by the teacher as indicators that one stage of the lesson has ended and that another one is beginning and may be a feature of all spoken discourse. "Framing moves" are frequently followed by "focusing moves", which are used to talk about the discourse, to inform the pupils what is going to happen or what has happened. "Opening moves" cause others to participate in an exchange. They consist of an utterance not structurally dependent on a previous turn. "Opening" and "answering" moves are complementary as "the type of answering move is pre-determined because its function is to be an appropriate response in terms laid down by the opening move" (Sinclair & Coulthard, 1975, p. 45). "Answering moves" usually consist of a maximum of three elements: pre-head, head and post-head. On many occasions only the head occurs. "Follow-up moves" are intended to let pupils know how well they have performed. "Follow-up has a three-term structure, pre-head, head, post-head, realised by accept, evaluate and comment respectively" (Sinclair and Coulhard, 1975:45) . Sinclair and Coulthard's (1975) tripartite classification of teaching exchanges moves will be referred to in chapter 6 within the notion of "good listenership" (McCarthy, 2002).

In a diachronic investigation of spoken interaction, Culpeper and Kytö (2010) summarise the prototypical characteristics of spoken conversation and detail characteristic grammatical

features[4] . Culpeper and Kytö (2010, p. 92) address six dimensions: production/reception, function, interactivity, sharedness, restrictions on format and transmission. "Production/reception" refers to the variability in language produced, and whether it is produced for one individual and by one individual. Individual characteristics include more vernacular or non-standard grammar (e.g. "*yous"*, "*ain't"*, and "*them shoes").* "Function" relates to the expression of speaker attitudes/states, and affirmation of the channel and social relations. Expressive, phatic features include more politeness formulae (e.g. "*thanks"*, "*sorry"*, "*could you"),* familiarising vocatives (e.g. "*mum", "guys", "mate"*), interjections (e.g. "*oh", "ah", "ha"*), expletives, exclamations (e.g. "*what a load of rubbish").* "Interactivity" addresses the interconnections linking utterances, i.e. forming exchanges and the speed of exchanges. Dialogue (rapid exchanges) include more second person pronouns, vocatives, questions, imperatives, negation, certain discourse markers (e.g. "*well"*, "*you know"),* inserts (e.g. "*yes","ok","alright","oh","huh")* and certain greetings.

"Sharedness" refers to the similarity of potential access to physical and temporal contextual features, to co-text and to background knowledge. Shared situation features include less elaboration (e.g. low phrase length), fewer content words, more pro-forms (e.g. personal pronouns), "*do*" as pro-verb, grammatical ellipsis (e.g. "*you okay"),* non-clausal material and inserts. "Restrictions on format" comprise the degree to which linguist forms, topic and turn-taking follow a pre-determined course. Freely developed spontaneous features include low token-type ratio (e.g. more repetition) and more pre-fabricated units or lexical bundles than academic writing. "Transmission" relates to the durability of linguistic substance, ability to back-track and plan in advance. Real-time processing features include low token-type ratio, less complex lexis and more normal dysfluencies (e.g. hesitation, pauses, repeats, false starts, grammatical reduction, grammatical dislocation[5]).

Culpeper and Kytö (2010, p. 93) indicate that the grammar of spoken conversation is linear as it proceeds incrementally, and dynamic as it takes place in real time. Leech in Biber et al. (1999, pp. 1066-1067) suggests three principles underlying the online production of spoken conversation, which account for many of the characteristics of the grammar of conversation:

[4] The characteristic grammatical features are mostly drawn from Leech (2000) and Biber et al. (1999) corpus-based investigations and are not necessarily exclusive to spoken conversation.
[5] Left dislocation refers to the phenomenon where items semantically co-referential with the subject or object of the class are positioned before the subject (Carter and McCarthy, 1999).

1. Keep talking: one cannot simply stop but can gain more time by hesitating, restarting, or one can simply give up the conversational floor.
2. Limited planning ahead: working memory is limited, perhaps to seven random items, thus, structures are usually less elaborate at the beginning or in the middle of a clause.
3. Qualification of what has been said: considering the above two factors, it may be necessary to retrospectively modify the message by adding on elements which, in other circumstances, might have been integrated earlier.

Furthermore, Culpeper and Kytö (2010, pp. 93-94) detail two aspects of the grammar of conversation: specific multi-word combinations or lexical bundles and specific structural patterns. "Lexical bundles" are relatively stable combinations of words which are not ordered by a set of abstract grammatical categories but are inseparable from grammar, i.e. they are the grammar. Biber et al. (1999, p. 994) indicate the most frequent three-word lexical bundles in conversation: "*I don't know"* (over 1,000 occurrences); "*I don't think"*, "*Do you want"* (over 400 occurrences); "*I don't want", "don't want to", "don't know what", "and I said', "I said to", "I want to", "you want to", "you have to", "do you know", "you know what"," have you got", "what do you", "I mean I", "have a look"* (over 200 occurrences).

In relation to particular structural elements, Culpeper and Kytö (2010, pp. 96-97) illustrate the following patterns associated with spoken conversation:

1. Prefaces

a) Discourse markers and other prefatory expressions:

- ***"You know** I never did get to spin".*
- ***"I mean** are these the same, these are the same?"*

b) Overtures:

- ***"I'll tell you what** I've just had a thought".*
- ***"You mean to say**, we're paying two and a half thousand pounds worth of repairs, I says, and they're not done?"*

c) Fronting:

- ***"Car number I remember** more by letters than the numbers".*
- *"And the numbers, **the numbers I don't remember** well. Some of them I do".*

d) Noun-phrase prefaces co-referential to pronouns

- ***"This little shop- it**'s lovely".*

- ***"North and south London they****'re two different worlds".*

2. Bodies

a) Extending:

- "*She doesn't like people smoking in her house –* ***because she says I don't want my bedroom smelling of smoke, her sheets – and everything else*****"**.

3. Tails[6]

a) Tag questions, comment clauses and retrospective hedges:

- "*Well, that little girl's cute* ***isn't she***?"
- "*Mm I wouldn't go into Amanda Close* ***I don't think*****"**.
- "*And it was her second car that she'd ever had* ***sort of thing*****"**.

b) Noun phrase tags:

- **"*That****'ll be a bit crispy,* ***that bit*****"**.
- "*Oh I reckon* ***they****'re lovely. I really do* ***whippets*****"**.

c) Retrospective elaboration or reinforcement:

- "*I don't care about the work and them being in a muddle,* ***no not at all*****"**.
- "*I mean, she* ***never*** *liked that car.* ***Ever*****".**

To sum up, in this sub-section I have presented Brown and Yule's (1983) view on the different demands spoken and written language make on language producers and highlighted the further challenges faced by speakers in face-to-face conversation due to its real time nature. Secondly, I have summarised Sinclair and Coulthard's (1975) account of the structure and classes of conversational moves, which can be applied to different conversational contexts. Thirdly, I have described Culpeper and Kytö's (2010) six-dimension model of prototypical characteristics of spoken conversations, the three principles underlying the online production of spoken conversation (keep talking, limited planning ahead and qualification of what has been said) and aspects of the grammar of conversation, namely lexical bundles and specific structural patterns.

In Culpeper and Kytö's (2010) characterisation of particular structural elements, one can notice that the discourse markers "*You know*" and "*I mean*" are also referred to as prefatory expressions. The term "discourse marker" is widely used to refer to different linguistic forms which perform different pragmatic functions. In order to distinguish what a discourse marker

[6] Tails are optional slots which are placed after the body of the message.

is from what it is not, I first address Halliday and Hasan's (1976) seminal treatment of cohesion in which they detail the cohesive relations which, in their view, enable a text be identified as such. This theoretical section prepares the ground for the characterisation of discourse markers in the subsequent sub-section.

1.1.2. Cohesion

The relationship between cohesion and coherence and their role in discourse are key topics within linguistic approaches to discourse. In their seminal treatment of cohesion, Halliday and Hasan (1976) present cohesive features which, in their view, can identify a text as such. They assume the principle that a text may be spoken or written, in prose or verse, a dialogue or a monologue. It can be anything from a single proverb to a whole play. In a narrower sense, a text is a semantic unit that conveys meaning. "A text does not consist of sentences, it is realised by, and encoded in, sentences" (Halliday & Hasan, 1976, p.2). They claim that a text has texture, a property that determines whether or not a set of sentences constitute a text. This texture is provided by the cohesive relationships within a text, which are set up where the interpretation of some element in the discourse is dependent on that of another. Halliday and Hasan (1976, p. 4) define cohesion as "the relations of meaning that exist within the text and that define it as a text". Such cohesive relationships are provided by means of reference, substitution, ellipsis, conjunction and lexical cohesion.

With regards to reference, Halliday and Hasan (1976) explain that certain items in every language have the property of reference since they are not interpreted semantically on their own but make reference to something else for their interpretation. This something else may lie within the text or outside the text, i.e. in the environment. Thus, there are two kinds of reference:

1. **Exophora**: it does not name anything but signals that reference must be made to some person or object in the environment. For instance: "*Look at **her**! Isn't **she** lovely*?" The pronouns "*her"* and "*she"* are used exophorically as they point to a person whose presence or image is only visible in the environment;
2. **Endophora:** reference must be made to something within the text. It can be divided into two sub-categories:

a) **Anaphora:** reference must be made to an element previously presented in the text. It is the most common form of reference in English and in many other languages. For instance: "*Mary was looking for you. Did you manage to speak to* ***her***?" The pronoun *"her"* is used anaphorically as it refers back to *"Mary";*

b) **Cataphora:** reference must be made to an element yet not presented in the text. For instance: "*Everyone had been waiting for* ***her****. Ladies had put on their best outfits to meet* ***her*** *in person. Suddenly, someone shouted:* ***The Queen*** *has arrived!*" The pronoun *"her"* is used cataphorically as it refers forwards to *"the Queen"*.

A second type of cohesive relation is achieved through ellipsis and substitution. Halliday (1985) explains that both ellipsis and substitution are variants of the same kind of cohesive relation. "Ellipsis" is a form of anaphoric cohesion where one presupposes something by means of what is left out (Halliday, 1985, p. 316). It contributes to the semantic structure of the discourse and sets up a relationship that is lexicogrammatical, i.e. a relationship in the wording rather than directly in the meaning. For instance:

a. "*Why didn't you use a credit card*?"
b. "*I hadn't got* ***any***".

In the exchange above, one needs to presuppose that the modifier "*any*" refers back to the noun "*credit card*", being an elliptical form of the noun phrase "*any credit card*".

When an explicit indication is given that something is omitted, by the use of a substituted form, this is called "substitution". For instance:

a. "*I've lost my* ***umbrella***".
b. "*Get a new* ***one***".

In the above exchange, the pronoun "*one*" has substituted the noun "*umbrella*" so as to avoid repetition of the noun.

Thirdly, explicitly marked cohesive relationships can be indicated by formal markers, which relate "what is about to be said" to "what has been said before". Halliday and Hasan (1976, pp. 238-239) categorise conjunctions into four types[7]:

1. **Additive:** "*For the whole day he climbed up the steep mountainside, almost without stopping.* ***And*** *in all this time he met no one.*"

[7] For a full account of the conjunctions comprising the above four types and their internal and external functioning, please refer to Halliday and Hasan (1976, pp. 241-267).

2. **Adversative:** "***Yet*** *he was hardly aware of being tired.*"
3. **Casual**: "***So*** *by night the valley was far below him.*"
4. **Temporal:** "***Then****, as dusk fell, he sat down to rest*."

As for lexical relationships, Halliday (1985) indicates that lexical cohesion comes about through the selection of lexical items which are related in some way to those which have previously appeared in the text. Examples of lexical cohesion include:

a) **Repetition of items:** "*Algy met* ***a bear***. ***The bear*** *was bulgy*."
b) **Synonymy:** noise/ sound
c) **Hyponymy (specific to general):** python/snake
d) **Meronymy (part to whole):** bush/hedge
e) **Antonymy:** asleep/woke
f) **Collocation (tendency of items to co-occur):** snow/ ice

Brown and Yule (1983) revisit Halliday and Hasan's treatment of cohesion. The former admit that most texts will indeed reveal some cohesive structuring. However, Brown and Yule pose three very pertinent and thought-provoking questions:

1) Is such cohesion necessary for the identification of a text?
2) Is such cohesion sufficient to guarantee identification as a text?
3) If a text is identified by these criteria, will they guarantee textual coherence?

Concerning the first question, Halliday and Hasan (1976) acknowledge that the notion of cohesion needs to be supplemented by the notion of register, which refers to appropriateness to a particular context of situation. However, they claim that the presence of some cohesive relationships is essential for a text to be identified as such:

> The concept of cohesion accounts for essential semantic relations whereby any passage of speech and writing is enabled to function as a text. We can systematise this concept by classifying it into a small number of distinct categories…Each of these categories is represented in the text by particular features…which have in common the property of signalling that the interpretation of the passage in question depends on something else. If that "something else" is verbally explicit, then there is cohesion. There are, of course, other types of semantic relations associated with a text which are not embodied in this concept; but the one that is does embody is in some ways the most important, since it is common to text of every kind and is, in fact, what makes a text a text.(Halliday & Hasan,1976, p.13)

For Brown and Yule (1983), the above quotation fails to draw a distinction between "meaning relations", which hold between items in a text, and "the explicit expression" of those meaning relations within a text. The authors question whether the explicit realisation of such relations is required to identify a text as a text. For Brown and Yule (1983), Halliday and Hasan (1976) seem to be talking of verbal elements that appear in the verbal record, not of the underlying semantic relations.

In relation to textual coherence, Halliday and Hasan (1976, p. 23) indicate that cohesion and register together effectively define a text. They add that a text is a passage of discourse which is coherent in two dimensions. A text is coherent with respect to the context in the situation, i.e. it is consistent in register. A text is also coherent with respect to itself, .i.e. it is cohesive. Neither of these conditions operates in isolation nor does one necessarily entail the other. For instance, it is possible to construct passages that appear to hang together in the situational-semantic sense but lack cohesion and, therefore, fail as texts. Similarly, it is also possible to construct beautifully cohesive passages that lack consistency of register and thus, fail as texts too.

Conversely, Brown and Yule (1983, p. 196) argue that it is common to find texts, in the sense of contiguous sentences[8] which we readily co-interpret, which display few, if any, explicit markers of cohesive relations:

1) A: "*There's the door bell*".
   B: "*I am in the bath*".
2) Thank you for your comments about voicing. I will eventually get back to that lesson. (Beginning of a letter).

In the above examples, there is no explicit marker of cohesive relation between the first and the second sentence of each sequence. Nevertheless, we will assume that the pair of sentences constitutes a text and will interpret the second sentence in the light of the first one. We will assume that there are semantic relations between the sentences, in the absence of any explicit assertion that there is such a relationship (Brown & Yule, 1983, p. 196). Thus, the authors claim that texture, in the sense of explicit realisation of semantic relations, is not criterial to the identification and co-interpretation of texts.

[8] Contiguous sentences may refer to adjacency pairs as in the first example. Pair adjacency will be addressed under Conversation Analysis in the sub-section 1.3.2.

As for the second and third questions, Brown and Yule (1983) advocate that formal cohesion does not guarantee identification of a text or textual coherence. In their view, hearers and readers do not depend upon formal markers in order to identify a text as a text. "Where language occurs contiguously in time and space, we attempt to co-interpret. However, we are constrained in these attempts very powerfully by the conventional presentation of the text" (Brown & Yule, 1983, p.198). For instance, when we read a newspaper, we do not simply read across the horizontal line for the processing of the written text but take into account other elements like columns, pictures, the type of print, etc. Likewise, when we listen to a person speaking, we also consider the voice quality, intonation and pauses in order to process spoken discourse. Thus, Brown and Yule (1983, p. 199) conclude that "texts are what hearers and readers treat as texts".

Halliday and Hasan's (1976) view that cohesion is a determinant of textness has been disputed by other researchers. Nonetheless, their description of cohesive relations, be them grammatical or lexical, has provided us with a clear understanding of how parts of discourse can effectively hang together. Similarly, Brown and Yule (1983) highlight the role of contiguity in the identification of texts. Their argument strengthens the notion of coherence in the interpretation of discourse. In the next sub-section, I focus on the role of discourse markers in the organisation and in the management of spoken discourse, namely face-to-face conversations.

1.1.3. Discourse markers

In the previous section, I briefly presented Halliday and Hasan's (1976) taxomony of conjunctions. Halliday and Hasan (1976) argue that conjunctive elements are cohesive not in themselves but indirectly, through their specific meanings. Although they are not primarily used for reaching out into the preceding or forthcoming texts, they express certain meanings which presuppose the presence of other elements in the discourse. In addition to the additive, adversative, casual and temporal conjunctive categories, Halliday and Hasan (1976, p. 267) describe other conjunctive elements that they name "continuatives".

Continuatives are a number of individual items which do not fit into the previous taxonomy of conjunctions, being more of a miscellaneous type, yet they are used with a cohesive force in

the text. With regards to prosody, when functioning cohesively, these elements are reduced forms (i.e. unaccented and with reduced vowel values) of items which co-occur whereas when being used non-cohesively, they are uttered as full forms. Halliday and Hasan (1976, pp. 268-271) discuss this prosodic (and pragmatic) distinction by comparing different uses of some of these elements and exemplifying their cohesive function:

1. Now:

a) Tonic: deictic[9] expression

b) Cohesive: it marks the opening of a new incident in a story, a new point, etc.

*"**Now** what would you like, dear?"*

2. Of course

a) Tonic: meaning = *you should have known that already.*

b) Cohesive: meaning = I accept the fact; if used rhetorically, it means *you must accept the fact.*

*"They were going to come to the meeting. **Of course** they may have changed their minds".*

3. Well

a) Tonic: meaning= *I acknowledge the question and will give it a considered answer*; often therefore amounting to no more than a hesitation noise: *I'm thinking about it*.

b) Cohesive: it serves to indicate that what follows is in fact a response to what has preceded; if used in a continuum by the same speaker, it introduces an explanatory comment.

A: "*Do you I look pale?"*

B: "***Well**- yes- a little".*

4. Anyway

a) Tonic: dismissive adversative meaning = *no matter under which, or what, circumstances*.

b) Cohesive: meaning= *let's get on with the job*.

[9] "Deixis" is a technical term borrowed from the Greek word, which means pointing or indicating via language. It refers to "the way languages encode or grammaticalise features of the context of the utterance or speech event", and thus, also refers to "ways in which the interpretation of utterances depends on the analysis of that context of utterance" (Levinson, 1983, p. 54). The linguistic units or morphemes used to accompany this "pointing" are called "deictic expressions". They are among the first forms young children utter and allow speakers to locate aspects such as "who", "where" and "when" during the ongoing verbal exchange. Thus, there are three traditional categories of deixis: person, place and time.

*"They changed over to a most peculiar kind of train which you don't see now. I've forgotten what it is called. Was it called a "steam coach"? **Anyway** it was just one coach but it ran on steam and it made a funny noise".*

5. Surely

a) Tonic: it invites the hearer to assent to the proposition being enunciated.

b) Cohesive: meaning = *am I right in my understanding of what has just been said*; sometimes specifically *you can't have meant …*?

*"They'll think you're serious. Nobody could be so stupid as to think that, **surely**".*

6. After all

a) Tonic: meaning = *after everything relevant has been considered, what remains is…*

b) Cohesive: meaning = *what I have just said is reasonable, when everything is taken into account.*

"*You needn't apologise. **After all** nobody could have known what would happen".*

Basically, on describing continuatives and their cohesive uses, Halliday and Hasan (1976) seemed to be talking about "discourse markers". Jucker and Ziv (1998) indicate that the notion of "discourse markers" is rather fuzzy as there is no generally agreed upon definition of the term. Discourse markers are also referred to as pragmatic markers, discourse particles, pragmatic particles, pragmatic expressions and connectives. This diversity reflects the wide range of linguistic approaches employed in their investigation and the multiplicity of functions these are said to fulfil. Such functions include: discourse connectors, turn-takers, confirmation-seekers, intimacy signals, topic-switchers, hesitation markers, boundary markers, fillers, prompters, repair markers, attitude markers and hedging devices.

Jucker and Ziv (1998) present two characterisations of discourse markers. For Hölker (1991, pp. 78-79), discourse markers or pragmatic markers have four basic features: they do not affect the truth conditions of an utterance; they do not add anything to the propositional content of an utterance; they are related to the speech situation and not to the situation talked about; they have an emotive, expressive function rather than a referential, denotative or cognitive one. For Brinton (1996, pp. 33-35), pragmatic markers display the following list of characteristics:

a) They are predominantly a feature of oral rather than of written discourse;
b) They appear with high frequency in oral discourse, sometimes with more than one occurrence in a single sentence;
c) Because of their frequency and oral nature, they are stylistically stigmatised and negatively evaluated, especially in written or formal discourse;
d) They are short items, often phonologically reduced or unstressed;
e) They form a separate tone group with falling-rising and rising intonation;
f) They are said to be restricted to sentence-initial position though studies reveal that they frequently appear sentence medially and finally as well;
g) They are considered to have little or no propositional meaning;
h) They occur either outside the syntactic structure or loosely attached to it and thus have no clear grammatical function;
i) They seem to be optional rather than obligatory features;
j) They are marginal forms;
k) They may be multifunctional, operating on local and global levels;
l) They are said to be more characteristic of women's speech rather than of men's.

Jucker and Ziv (1998) clarify that Brinton uses a number of hedges when describing the above features such as "*predominantly*", "*often*" and "*sometimes*". Despite its tentative and sketchy nature, Jucker and Ziv (1998) acknowledge that Brinton's list suggests a range of features that discourse markers may display. Jucker and Ziv (1998) also add that very few, if any, discourse markers will manifest all the above features. However, the higher the number of features, the more prototypical the specific marker can be regarded. Furthermore, the phonological, syntactic and semantic features are decisive in the assessment of discourse markers.

Culpeper and Kytö (2010) draw a distinction between pragmatic markers and discourse markers. In their view, "pragmatic markers have little or no propositional meaning but tell us about the pragmatic relations between the speaker, their message(s) and its context" (Culpeper & Kytö, 2010, p. 361). Likewise, discourse markers have the additional feature of mediating discourse, i.e. they specifically mediate between one speaker's utterance and another. As for structural patterns, pragmatic markers are usually short items and do not form a single category. They belong to marginal categories that operate outside the core syntax such as

adverbs, adverbials, conjunctions and parenthetical clauses. Discourse markers have the additional feature that they are often in initial position with respect to the main clause, and in spoken conversation, form a separate tone group. For Fraser (1996, p. 186), discourse markers are a subtype of pragmatic markers in the sense that they signal "the relationship of the basic message to the foregoing discourse".

Returning to our discussion on discourse markers, Carter and McCarthy (2006) define discourse markers as words and phrases whose function is to link segments of discourse in ways that reflect choices of monitoring, organisation and management by speakers. Based on corpus-evidence, they indicate that the most common discourse markers in spoken English consist of single words items. With regards to their functions, speakers use discourse markers to manage the discourse in terms of launching and concluding topics, opening, concluding or temporarily closing a conversation, re-opening previously closed or interrupted conversations. Among the most common and frequent markers that facilitate openings and closings are single words like "*so"*, "*(all) right"*, "*right then"*, "*now"*, "*good"*, "*well"*, "*okay (then)"* , "*anyway"* and "*fine"* (Carter & McCarthy, 2006, p. 214). Secondly, discourse markers signal relationships of sequence by indicating explicitly the order in which things occur or the organisation of the different segments. Among the words and phrases that perform this sequencing function in spoken language are (Carter & McCarthy, 2006, p. 216): "*and"*, "*and then"*, "*finally"*, "*first(ly)"*, "*for a start"*, "*going back to"*, "*in general"*, "*in the end"*, "*in the first place"*, "*last of all"*, "*lastly"*, "*next"*, "*on top of that"*, "*second (ly*)", "*so"*, "*there again"*, "*third(ly)"*, "*to sum up"* and "*what's more"*.

A few discourse markers also allow speakers to monitor and manage the ongoing discourse by commenting explicitly on the process of talking itself. Speakers can indicate that they have not selected the most appropriate way of expressing things and that they are adding to or refining what they say. Among reformulation markers are (Carter & McCarthy, 2006, pp. 220-221): "*as I was saying"*, "*as it were"*, "*I mean"*, "*if you like"*, "*in a manner of speaking"*, "*in other words"*, "*not to say"*, "*or rather"*, "*so to speak"*, "*strictly speaking"*, "*that's to say"*, "*to put it another way"*, "*to put it bluntly/mildly"* and "*well"*. Speakers can also signal that they are sensitive to listeners' needs and that they are monitoring the state of shared knowledge through the use of markers such as "*you know"* and "*you see"*. Similarly,

discourse markers also enable listeners to indicate their involvement[10] with what is being said and to manage their own responses such as "*all right"*, "*I see"*, "*good"*, "*great'*, "*fine"* and "*okay"*.

Lastly, discourse markers can be used to mark speakers' stance or attitude towards the message. Among the most common and frequent stance markers are (Carter & McCarthy, 2006, p. 222): "*actually"*, "*admittedly"*, "*amazingly"*, "*basically"*, "*certainly"*, "*clearly"*, "*confidentially"*, "*doubtless"*, "*essentially"*, "*frankly"*, "*to be frank"*, "*fortunately"*, "*honestly"*, "*to be honest"*, "*hopefully"*, "*ideally"*, "*if you ask me"*, "*I'm afraid"*, "*I must admit"*, "*I must say"*, "*I think"*, "*in fact"*, "*indeed"*, "*literally"*, "*naturally"*, "*no doubt"*, "*obviously"*, "*of course"*, "*predictably"*, "*putting* (*or to put*) *it mildly/bluntly"*, "(*quite*) *rightly"*, "*really"*, "*sadly"*, "*seriously"*, "(*I'm*) *sorry"*, "*strictly speaking"*, "*surprisingly"*, "*thankfully"*, "*to tell you the truth"*, "*understandably"*, "*undoubtedly"* and "*unfortunately"*. In order to avoid sounding too blunt and assertive, speakers can also make use of discourse markers to hedge, i.e. to express different degrees of assertiveness. Among common hedges are (Carter & McCarthy, 2006, p. 223): "*apparently"*, "*arguably"*, "*by any chance"*, "*I think"*, "*just* (*about*)", "*kind of"*, "*like"*, "*maybe"*, "*perhaps"*, "*presumably"*, "*probably"*, "*roughly"*, "*sort of"* and "*surely"*. Interjections may also be used to mark positive or negative emotional reactions to what is being or has just been said or to something in the situation. They are especially common in spoken language and rare in written representations of speech. Among common interjections in spoken English are (Carter & McCarthy, 2006, pp. 224-225): "*aargh"*, "*bother"*, "*crikey"*, "*damn"*, "*god"*, "*goodness* (*me*)", "*gosh"*, "(*good*) *heavens"*, "*hooray"*, "*jeez"*, "*ooh"*, "*oh no"*, "*ooo"*, "*oops"*, "*ow"*, "*phew"*, "*poo"*, "*ugh"*, "*urgh"*, "*tut-tut"*, '*whoops"*, "*wow"*, "*yippee"* and "*yuk"*.

In this section, I have presented different characterisations of discourse markers. For Halliday and Hasan (1976), continuatives are linguistic elements which do not fit into their taxonomy of conjunctions but have a cohesive function. When functioning cohesively, these elements are uttered as reduced forms of items which co-occur. For Hölker (1991), discourse markers or pragmatic markers do not affect the truth conditions of an utterance; they do not add anything to the propositional content of an utterance; they are related to the speech situation and not to the situation talked about; they have an emotive, expressive function rather than a

[10] Discourse markers for expressing engaged listenership are addressed as non-minimal response tokens in chapter 6.

referential, denotative or cognitive one. For Brinton (1996), phonological and lexical, syntactic/textual and semantic features are decisive in determining whether words or phrases fit into the category of discourse markers. For Carter and McCarthy (2006), discourse markers are words and phrases whose function is to link segments of discourse in ways that reflect choices of monitoring, organisation and management by speakers. Based on corpus evidence, Carter and McCarthy (2006) detail the organisational and pragmatic functions of discourse markers.

The notion of "discourse marker" is closely related to discourse analysis as discourse markers are assumed to establish cohesive relations in spoken discourse. Similarly, the term "pragmatic marker" is usually associated with pragmatics. I follow Fraser's position (1996) that discourse markers are a sub-type of pragmatic markers used to mediate segments of discourse. In the next section, let us shift our focus to pragmatic theories that address the pragmatic relations between the speaker, the hearer, the meaning potential of an utterance and its context.

1.2. PRAGMATICS

Pragmatics is commonly associated with utterance or speaker meaning. But what are the elements and who are the agents that come into play when conveying, interpreting, negotiating or co-constructing meaning in interactional settings? Definitions of pragmatics stress different views and perspectives. For Levinson (1983), pragmatics is essentially concerned with inferences. When we make an inference, we arrive at a hypothesis, idea or judgement based on other knowledge, ideas or judgement. An inference is the reasoning which leads to a conclusion drawn from a premise. There are different types of inferences. On one hand, classical deductive inferences are forms of reasoning based on rules of formal logic. These inferences are necessarily valid, not defeasible[11] and can be demonstrated (Allwood et al., 1977, Sperber & Wilson, 1995). Semantic inferences relate to the decoding of utterances via the application of phonological, syntactic, morphological and lexical rules, and whose propositions[12] may or may not express truth conditions. Pragmatic inferences are based on the notion of implicatures proposed by Grice (1975), who emphasised the distinction between

[11] A defeasible inference can be cancelled by contextual features, false arguments, invalid argumentation and contradiction (Sperber & Wilson, 1995).
[12] A proposition is what a sentence says about the world (Allwood, Andersson & Dahl, 1977).

what words mean, what a speaker literally says when using them and what the communicative intention of the speaker is when using the words, which often goes beyond what is said.

From a broader perspective, Thomas (1995, p. 22) proposes a definition of pragmatics as "meaning in interaction" since meaning is not something inherent in words alone, nor is produced by speakers or hearers alone. "Making meaning is a dynamic process , involving the negotiation of meaning between speaker and hearer, the context of utterance (physical, social and linguistic) and the meaning potential of an utterance" (Thomas, 1995, p. 22). By the meaning potential of an utterance, the author refers to the possible interpretations of an utterance that are tied to its context. For example, when a speaker meets someone he or she has not seen for a long time in the street and utters "*How are things*", the hearer may interpret it as a purely phatic greeting or as a general question. It is unlikely that the hearer will interpret the utterance as a request to talk about his or her private life, bodily functions or health problems.

In their book "Pragmatics and the English Language*"*, Culpeper and Haugh (2014) address the notions of "inferences*"* and "meaning in interaction", among many others. They initially describe two views on the scope of Pragmatics. The narrow view originates from the works of early writers like Plato and Kant, the American philosophers Charles Peirce (1839-1914) and Charles Morris (1901-1979) and the German-American philosopher Rudolph Carnap (1891-1970). Based on Peirce and Carnap's work, Morris (1938) provided a point of departure for Pragmatics by proposing a three-way distinction for sign relationships:

a) **Syntax:** mono relationships (relationships between linguistic signs);
b) **Semantics**: dyadic relationships (relationships between linguistic signs and the things in the world that they designate);
c) **Pragmatics:** triadic relationship (relationships between linguistic signs, things they designate and their users/interpreters).

The above distinction signals that Pragmatics is the area that deals with context, interacting to a certain extent with syntax and semantics. For Culpeper and Haugh (2014), Morris took a micro view of context as he only mentioned users and interpreters, neglecting social relations and situations. Grice's Conversational Implicatures (1975) and Sperber and Wilson's Relevance Theory (1995) have their theoretical foundations in this micro view of context,

since they focus on users' intentions and on interpreters' inferences. Following from this, Pragmatics is often seen as another component in a theory of language, contributing to phonetics, phonology, morphology, grammar/syntax and semantics. This view of Pragmatics is traditionally identified as the Anglo-American view and addresses topics like reference, deixis, presupposition, speech acts, implicatures and inferencing.

The broad view, also referred to as the continental view, also includes the above topic areas, but encompasses much beyond them. Pragmatics is viewed as the superordinate field, with disciplines such as linguistics, sociology and psychology as sub-fields. Thus, pragmatics does not only add a contextual dimension to a theory of language but it is a "general cognitive, social, and cultural perspective on linguistic phenomena in relation to their usage in forms of behaviour" (Verschueren, 1999, p. 7). Culpeper and Haugh (2014, p. 7) point out that the previous quotation includes three important notions:

1. Pragmatics is not only sited within linguistics, but could equally be within cognitive, social and cultural fields of study;
2. Pragmatics only looks at linguistic phenomena in actual usage, rather than at linguistic phenomena per se;
3. The scope of pragmatic analysis includes behaviour, i.e. what people do, whether with language or with other means such as gestures, in social contexts.

In short, the broad view highlights a socio-cultural perspective on the functioning of language.

Culpeper and Haugh (2014, p. 7) explain that although both views seem to share an interest in cognition, their emphases differ. The narrow view focuses on the cognitive intentions and inferences involved in generating a speaker's meaning or re-constructing a hearer's understanding of it. The broad view would also encompass broader cognitive notions such as the way knowledge about situations, social institutions and cultures, among others, would influence or be influenced by language. Culpeper and Haugh (2014, p. 7) add that "one should not over-emphasise the differences between the Anglo-American and the Continental European views". They indicate that the topic "Politeness", for instance, can be addressed in both views since it aims at explaining not only some aspects of linguistic structure but also some aspects of social function and context.

Furthermore, many of the topics addressed in both narrow and broad views derive from three major pragmatic theories: Speech Acts (Austin, 1962, Searle, 1969, 1975), Conversational Implicatures (Grice, 1975) and Politeness (Brown & Levinson, 1978, 1987). Austin was intrigued by the way speakers use words to do different things. Austin's ideas were further exploited by his student, John R. Searle (1969), who developed the Speech Act Theory by creating a taxonomy of speech acts. Similarly, Grice's Conversational Implicatures Theory is based on his distinction between what someone says and what someone implicates by uttering a sentence. Brown and Levinson's Politeness is based on the principle that human communication is governed by politeness rules which are basic to the production of social order and work as a pre-condition of human cooperation. For the purposes of this thesis, Speech Acts will be discussed as a separate sub-section whereas Implicatures and Politeness will be addressed within the same sub-section.

1.2.1. Speech acts

In the late thirties, Austin challenged the view of language that placed truth conditions as central to language understanding. For the philosopher, when people want to express themselves, they may produce utterances with propositional content, which can be true or false in the real world, such as "*The new Pope is Argentinean*". However, not all utterances fall into this category. Some ordinary language declarative sentences are not necessarily used with any intention of making true or false statements about the world. (Austin, 1962, p. 5):

1. "*I name this ship Queen Elisabeth*" – as uttered when smashing a bottle against the stem;
2. "*I give and bequeath my watch to my brother*" – as occurring in a will;
3. "*I bet you six pence that it will rain tomorrow*".

According to Austin (1962), the above sentences are not used just to say things but rather to do things, i.e. to perform actions. These sentences cannot be assessed in terms of truth or falsity. Therefore, Austin initially coined the term "performatives" to refer to these peculiar sentences and contrasted them with "constatives", which refer to sentences, assertions and utterances carrying truth conditions. However, Austin (1962, pp. 14-15) pointed out that performatives depend on certain conditions to be successful or felicitous:

A. (i) There must be a conventional procedure having a conventional effect

(ii) The circumstances and persons must be appropriate

B. The procedure must be executed (i) correctly, (ii) completely

C. Often

(i) The persons must have the requisite thoughts, feelings and intentions and

(ii) If consequent conduct is specified, then the relevant parties must do it.

Austin also referred to verbs explicitly used to perform actions as "performative verbs" as in the examples below:

"*I **pronounce** you man and wife*".

"*I **apologise** for any inconvenience this may have caused*".

Culpeper and Haugh (2014, p. 156) clarify that the performativity of a performative verb does not depend entirely on its semantics but it is sensitive to the linguistic and non-linguistic context. For instance, in the sentence "*I apologise all the time*", the verb "*apologise*" is used descriptively, i.e. it describes what the speaker does all the time instead of being used performatively.

Thomas (1995, pp. 33-44) categorises Austin's performatives as follows:

a) Metalinguistic performatives: verbs which are self-referential (referring to what the speaker of the utterance is doing), self-verifying (containing their own truth condition) and non-falsifiable (they can never be untrue). Examples in English include: "*I say*", "*I protest*", "*I object*" and so forth;

b) Ritual performatives: they are highly cultural dependent and associated with highly institutionalised contexts. For instance: "*I sentence you to ten years…*" and "*I name this ship…*";

c) Collaborative performatives: verbs which require the collaboration or particular uptake of another person. For example: "*I bet you ten pounds…*" and "*I challenge you …*";

d) Group performatives: they are either commonly or necessarily produced by more than one person and may fall into any of the above categories. For instance: "*We do not judge you to be guilty of professional misconduct*" (Thomas, 1995, p. 41).

At a later stage, Austin moved away from the dichotomy constative-performative and argued that utterances do not require a performative verb to perform an act. Let us examine the examples below:

a) "*I promise to finish that paper by Tuesday*".
b) "*I'll finish that paper by Tuesday*".

Both utterances involve the act of promising. While utterance (a) is an "explicit performative" as it contains a performative verb in its composition, utterance (b) is an "implicit performative" as it performs the act of promising without the presence of a performative verb.

Following from this, Austin (1975) proposed a three-fold distinction among related acts which take place every time an utterance is produced:

1. **Locutionary act:** the basic act of saying something; uttering words with certain sense and reference;
2. **Illocutionary act**: utterances which carry a certain conventional force such as a "statement", a "promise", a "warning", etc. The illocutionary act is performed via the communicative force of an utterance, i.e. the illocutionary force;
3. **Perlocutionary act:** the effects speakers produce upon the feelings and actions of their interlocutors, which are special to the circumstances of the utterance.

Culpeper and Haugh (2014, p. 160) point out that Austin's three-fold distinction separates meaning, traditionally defined in terms of sense and reference, from performing a function, as exemplified below:

Context: in a hot classroom, a teacher addresses a student sitting next to the window:

1. Locutionary act: the actual uttering of "*It's hot in here*";
2. Locutionary act: performing an act of request through this utterance;
3. Perlocutionary act: getting the student to open the window.

Culpeper and Haugh (2014) clarify that, although speech act theory comprises all three facets, the notion of "speech act" is virtually synonymous with illocutionary act or illocutionary force in further developments of the theory.

John Searle (1969) developed Austin's theory further by formalising the notion of felicity conditions, by creating a taxonomy of illocutionary acts and the concept of illocutionary force indicating devices and by addressing indirect speech acts. Initially, Searle (1969) categorises

the constitutive rules underlying the successful performance of illocutionary acts, namely felicity conditions, as follows:

1. **Propositional content conditions:** rules which put conditions on the propositional content of some illocutionary acts, e.g. in the case of "a promise" or "a warning", the content must be about a future event;
2. **Preparatory conditions:** they tell us what the speaker will imply in the performance of the illocutionary acts. In the case of "a promise", for instance, firstly, the event will not happen by itself, and secondly, the event will have a beneficial effect;
3. **Sincerity conditions:** they inform us of what psychological state the speaker expresses to be in, e.g. for "a promise", the speaker genuinely intends to carry out the future action;
4. **Essential Conditions:** they indicate what the action consists in essentially. By the act of uttering "a promise", for example, the speaker thereby intends to create an obligation to carry out the action as promised.

Secondly, Searle (1976) proposes a taxonomy of illocutionary acts into five mutually exclusive and jointly exhaustive classes**:**

1. **Representative or Assertive:** statements which commit the speaker to the truth of the assumption expressed; paradigm cases: asserting, concluding. Example: *"It's raining".*
2. **Directive:** speech acts speakers use to get someone else to do something; paradigm cases: requesting, questioning, commanding. Example: *"Close the door!"*
3. **Commissive:** they commit the speaker to the performance of a future action; paradigm cases: offering, threatening, promising. Example: *"I'll finish the paper by tomorrow"*.
4. **Expressive:** they convey the speaker's emotional attitude to the assumption expressed; paradigm cases: thanking, welcoming, congratulating, apologising. Example: *"I am sorry I did not get the chance to say goodbye to you"*.
5. **Declarative:** statements which effect immediate changes in the institutional state of affairs and which tend to rely on elaborate extra-linguistic institutions; paradigm cases: excommunicating, declaring war, firing from employment, christening. Example: *"I baptise you in the name of the Father, the Son and the Holy Spirit!"*

Culpeper and Haugh (2014) explain that other speech-act theorists have put their efforts into classifying speech acts, for instance, Bach and Harnish (1979), Ballmer and Brennenstuhl (1981) and Wierzbicka (1987). However, no definite taxonomy has emerged so far, which stresses the difficulty of classifying language functions of any type. Yet, such taxonomies have served for some analytical purposes.

A third contribution to speech act theory is Searle's notion of illocutionary force indicating devices (IFIDs). Searle (1969) claims that the illocutionary force indicator shows how the proposition is to be taken, i.e., what illocutionary force the utterance is to have. "Illocutionary force indicating devices in English include at least: word order, stress, intonation contour, punctuation, the mood of the verb, and the so-called performative verbs" (Searle, 1969, p. 30). He adds that speakers can indicate the type of illocutionary act being performed by beginning the sentence with performative verbs, for instance, "*I apologise*" and "*I warn*". Nonetheless, in actual speech situations, the context will help to establish what the illocutionary force of the utterance is, without the need for invoking the appropriate explicit illocutionary force indicator. Similarly, Culpeper and Haugh (2014, p. 168) stress that even utterances that contain IFIDs do not necessarily convey a particular illocutionary force, which may be dependent on the rest of the discourse and context. For instance, the utterance, "*I promise I'll withhold your pocket mone*y" is not a promise but a threat. Culpeper and Haugh (2014) refer to this example to illustrate the mismatch between sentence type and speech act correspondence.

In English, there are three basic sentence types associated with three speech acts:
a) Declarative/ assertion: "*I've finished my dinner*".
b) Interrogative/question/enquire: "*Have you finished your dinner*?"
c) Imperative/command: "*Finish your dinner!*"

As we can observe in the above examples, there is a correspondence between sentence type and speech act. Nevertheless, a particular speech act may be performed through different forms. Requests, for example, can appear in three sentence types: "*Open the window*" (imperative), "*Could you open the window, please?*"(interrogative) and "*It's hot in here*" (declarative). Requests also differ in terms of directness. While the imperative form is more direct, the interrogative form is indirect and the declarative form is basically a hint. Culpeper

and Haugh (2014) add that the mismatch between sentence types and speech acts is quite frequent in present-day English and indicate that the second example characterises Searle's (1975, p. 30) classic definition of indirect speech acts, which are "cases in which one illocutionary act is performed indirectly by way of performing another". At a later stage, Searle (1979) revisits his notion of indirect speech acts and argues that, in indirect speech acts, speakers communicate to the hearers more than they actually say by relying on mutual shared background knowledge and on hearers' inferential and rational abilities. Searle acknowledges that an account of the indirect part of indirect speech acts includes a theory of speech acts and certain general principles of cooperative conversation, some of which have been discussed by Grice (1975).

Culpeper and Haugh (2014) state that there is little agreement on the status of direct and indirect speech acts and how the latter work. However, they highlight the work of Blum-Kulka and her colleagues (1989) in their application of notions of indirectness to requests. Blum-Kulka, House and Kasper (1989, pp. 278-281) identify nine (in) directness strategy types:

Direct (impositives)

1. Mood derivable: "*Give me a hand*".
2. Performatives: "*Can you give me a hand?*"
3. Hedged performatives: "*Might I ask you to give me a hand?"*
4. Obligation statements: "*You must give me a hand*".
5. Want statements: "*I want you to give me a hand*".

Conventionally indirect

6. Suggestory formulae: "*Why don't you give me a hand?*"
7. Query preparatory: "*Can you give me a hand?*"

Non-conventionally indirect (hints)

8. Strong hints: "*This would be easier if someone gave me a hand*".
9. Mild hints: "*This is going to take me ages to do*".

The above categorisation illustrates the subtleties of meaning conveyed by form. Undoubtedly, mismatches between form and function are likely to pose problems to language learners when it comes to interpreting indirect speech acts. Corsetti (2009) investigated the comprehension and the production of speech acts by eight Brazilian learners of English, who

were taking a preparatory course for the IELTS examination, as part of a broader empirical project on pragmatic sub-competences. A pragmatic quiz was designed, which included two tasks intended to assess the understanding and the production of speech acts, among other tasks. In the first activity, learners were asked to match 14 utterances that depicted both direct and indirect speech acts to their illocutionary force. The second activity consisted of eight communicative situations to which learners were expected to write down how they would respond.

Task one highlighted learners' difficulty in recognising the illocutionary force of speech acts, particularly that of indirect ones. Four out of eight subjects were able to identify less than 43% of the illocutionary force of the 14 utterances whereas two subjects successfully recognised 67% of the illocutionary acts on average, and two subjects succeeded in identifying 85% of the illocutionary acts. Conversely, task 2 highlighted subjects' ability to produce illocutionary acts according to the situation, to the degree of formality and to the degree of imposition of face threatening acts[13]. Subjects' appropriate language usage ranged from 68% to 90%. Overall, the comparison between the results indicated that the interpretation of the illocutionary force of utterances posed a problem to 50% of the subjects whereas the production of illocutionary acts was successfully and appropriately carried out by 100% of the subjects. These findings[14] corroborated the assumption that Corsetti's learners had more difficulty interpreting rather than producing pragmatic meaning. After a three-month period in which learners were exposed to a series of strategy-based listening activities, students were tested again on their understanding of speech acts via a second pragmatic quiz. Results indicated a considerable improvement in their assignment of the illocutionary force of speech acts.

To sum up, speech acts play a vital role in the production and understanding of pragmatic meaning as they depict the functions of language usage. Culpeper and Haugh (2014) suggest that speech acts may be the most important theory in pragmatics. While Austin focused on language usage, specifically by detailing formal structures and examining the social contexts in which they were used, Searle addressed the notion of (in) directness, which has been

[13] According to Brown and Levinson (1987), face-threatening acts refer to acts produced by speakers which intrinsically threaten face, namely those acts that by their nature run contrary to the face wants of the addressee (H) and or of the speaker (S). To be detailed in the next sub-section 1.2.2.

[14] For more information on this empirical investigation and its results, please see Corsetti (2009, pp. 132-149).

largely investigated in current cross-cultural studies. Sperber and Wilson (1995, p. 244) state that "Speech Act Theory offers itself as a natural complement to Gricean pragmatics, dealing with the classification of speech-act in terms of both explicatures[15] and implicatures".

1.2.2. Implicatures and Politeness

The pragmatic phenomena addressed in this section, namely conventional and conversational implicatures, politeness and politeness strategies, are grounded on Grice's distinction between what someone says and what someone implicates when producing an utterance. Grice (1975) proposed some guidelines which underlie the efficient co-operative use of language and which jointly express a general co-operative principle. These guidelines are defined in terms of four conversational maxims (Grice, 1989, pp. 26-27):

The co-operative principle: Make your contribution such as required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

1. The Maxim of Quantity

(i) Make your contribution as informative as is required (for the current purposes of the exchange).

(ii) Do not make your contribution more informative than is required.

2. The Maxim of Quality

Try to make your contribution one that is true:

(i) Do not say what you believe to be false.

(ii) Do not say that for which you lack adequate evidence.

3. The Maxim of Relation (Relevance)

Be relevant.

4. The Maxim of Manner

[15] Carston (2004), a proponent of Relevance Theory, explains that the notion of "explicature" originated from the relevance framework, as a partner to the more common term "implicature". Although the term "explicature" may be related to the Gricean notion of "what is said", it also departs significantly from it given that an explicature "involves a considerable component of pragmatically derived meaning, in addition to the linguistically encoded meaning" (Carston, 2004, p. 3).

Be perspicuous:

(i) Avoid obscurity of expression.

(ii) Avoid ambiguity.

(iii) Be brief (avoid unnecessary prolixity).

(iv) Be orderly.

Speakers respecting the conversational maxims would speak sincerely, relevantly and clearly, while providing sufficient information. However, real life communication seldom meets these requirements. Grice (1989, p. 30) explained that a participant in a talk exchange may fail to fulfil a maxim in various ways:

a) "He may quietly and unostentatiously violate a maxim; if so, he may be liable to mislead";
b) "He may opt out from the operation both of the maxim and of the Co-operative principle; he may say, indicate, or allow it to become plain that he is unwilling to cooperate in the way the maxim requires";
c) "He may be faced by a clash";
d) "He may flout a maxim; that is, he may blatantly fail to fulfil it".

Speakers flout these maxims in the normal course of a conversation and, above all, this flouting of maxims indicates that a speaker is trying to say something else beyond the conventional meaning of the sentence uttered. In order to convey the implicit meaning of an utterance, speakers rely on a deeper level of co-operation that goes beyond sentence meaning. "Conversational implicatures" are then inferences which arise to preserve the assumption of co-operation. To illustrate a conversational implicature, let us examine the following example:

A: "Where's John?"
B: "There's a red BMW outside Joana's house".

If taken literally, (B)'s response fails to answer (A)'s question, flouting the Maxims of Quantity and Relevance. However, this apparent failure of co-operation indicates that (B) is relying on (A)'s co-operation to interpret the implicit meaning of the utterance: John has a red BMW, and therefore, may be in Joana's house. "Conventional implicatures", on the other

hand, are generated by the meaning of certain particles like the sentential connectors "*but*" and "*therefore*". Let us consider the following examples:

1. "He is an Italian, therefore he is loud".
2. "He is an Italian and he is loud".
3. "His being loud follows from his being Italian".

According to Grice, in (1) and (2) the speaker has said the same. The difference is that (1) entails (3), being a conventional implicature conveyed by the meaning of "*therefore*", and not by the flouting of the conversational maxims.

Grice (1975) also distinguished between kinds of conversational implicatures: generalised and particularised**.** In Grice's terms, "generalised conversational implicatures" (GCIs) arise without any particular context or special scenario being necessary whereas "particularised conversational implicatures" (PCIs) require such specific contexts. Levinson (2000) addresses the distinction between generalised and particular conversational implicatures by defending the existence of "preferred or default interpretations", which form the basis of a new radical theory of meaning. Levinson proposes a level of systematic pragmatic inference based on general expectations about how language is normally used rather than on direct computations about speaker-intentions. This third level of meaning is called "utterance-type-meaning".

GCIs are default inferences that capture our intuitions about a preferred or normal interpretation. In other words, these intuitions give rise to presumptions, default inferences, about both content and force. According to Levinson (2000, p. 23), it is at this intermediate level that speech acts, presuppositions, conventional implicatures, felicity conditions, conversational pre-sequences, and above all, generalised conversational implicatures operate.
In order to illustrate the distinction between the two levels of pragmatic inferences (generalised versus particularised conversational implicatures), Levinson (2000, p. 16) presents the following examples:

Example 1

A: "*What time is it?"*
B: "*Some of the guests are already leaving"*.
GCI= Not all the guests are already leaving.

PCI= It must be late.

Example 2

A: "*Where's John?"*

B: "*Some of the guests are already leaving"*.

GCI= Not all the guests are already leaving.

PCI= Perhaps John has already left.

Although the utterance-form *"Some of the guests are already leaving"* carries different particularised conversational implicatures (PCIs), which may be attributed to the Maxim of Relevance, there is a shared inference that "not all of the guests are in the process of leaving", which applies to both contexts. This preferred inference is obtained by the statement of the form: "**Some x are G**". In order to limit the range of possible default interpretations conveyed by a single utterance, Levinson relies on three heuristics, which are closely related to three of Grice's conversational maxims.

Another theory that is based on Grice's maxims is Brown and Levinson's Politeness (1978, 1987). Grice (1989, p. 28) acknowledged that "there are, of course, all sorts of maxims (aesthetic, social, or moral in character), such as "Be polite", that are normally observed by participants in talk exchanges". Politeness is a theory about social interaction since when people verbally interact, politeness phenomena are reflected in their linguistic behaviour. Brown and Levinson (1987, p. 55) claim that human communication is governed by politeness rules which account for "the linguistic minutiae of the utterances with which persons choose to express themselves in quite unrelated language and cultures". Although language usage may apparently seem nonsensical at times, there are some universals that characterise politeness and whose nature is rational. Thus, politeness is seen as a universal principle that is basic to the production of social order and a pre-condition of human cooperation.

Within the realm of politeness, Brown and Levinson (1987, pp. 61-62) present the following universal features shared by all competent adult members of a society:

1. Face: the public self-image that every member wants to claim for himself, which can be of two kinds:

(a) Negative face: freedom of action and freedom of imposition; the want of every competent member that his actions be unimpeded by others;

(b) Positive face: the positive consistent self-image or personality claimed by interactants, crucially including the desire that this self-image be appreciated and approved of; the want of every member that his wants be desired to at least some others;

2. Rationality: certain rational capacities, in particular consistent modes of reasoning from ends to means that will achieve those ends.

Considering the concepts of face and rationality mentioned above, Brown and Levinson (henceforth B&L) state that some acts produced by speakers intrinsically threaten face, namely those acts that by their nature run contrary to the face wants of the addressee (H) and or of the speaker (S). By "acts", B&L refer to what is intended to be done by verbal or non-verbal communication. "Face Threatening Acts" (FTAs) can be characterised as those which threaten negative face and those which threaten positive face. B&L advocate that when speakers perform FTAs, they use four super-strategies[16]: "bald on record", "positive politeness", "negative politeness" and "off record".

At one end of politeness, "bald-on-record strategies" include direct forms such as imperatives and basically refer to conforming to Grice's Conversational Maxims. In order to minimise a possible FTA, speakers may use mitigating devices. At the other end, "off-record strategies" rely on the flouting of particularised conversational maxims and encompass indirect uses of language and inferences, which need to be interpreted by addressees. "Positive politeness strategies" are oriented towards the positive face of (H) and may be used to convey common ground, and as a result, minimise potential FTAs. Conversely, "negative politeness strategies" are aimed at minimising the degree of imposition FTAs carry, respecting (H)'s freedom of action. According to B&L (1987), women have a tendency to use more elaborate positive politeness strategies than men in most cultures.

Recent developments in Politeness theory address its other side, namely Impoliteness. According to Culpeper (2011), Brown and Levinson (1978, 1987) focused on harmonious

[16] Politeness strategies were previously detailed and exemplied in Corsetti (2009, pp. 33-36).

interactions, ignoring impoliteness. Thus, impoliteness seemed to be regarded more as a pragmatic failure or anomalous behaviour. Nonetheless, Graig et al. (1986) introduce the notion of "face-attack" or "face aggravation" in relation to politeness theory for the first time. Culpeper (1996) then provides a comprehensive treatment of face-attack strategies. Culpeper et al. (2003) develop the theory further by examining how impoliteness can be both deployed and countered over sections of discourse longer than a single speech act. Following from this, Culpeper (2005) abandons B&L's (1987) distinction between positive and negative face and assesses interactions within context. Culpeper (2011) examines the various components that comprise or are connected with impoliteness attitudes, investigates the metadiscourse of impoliteness and discusses the functions of impoliteness events.

Concluding, in this section, I have detailed conventional and conversational implicatures, politeness and politeness strategies. Grice's unquestionable contribution to the study of utterance meaning via the notion of conversational implicatures remains unchallenged and fundamental to the underlying principles of contemporary pragmatic theories such as Generalised Conversational Implicatures (Levinson, 2000) and Brown and Levinson's Politeness Theory (1978, 1987). Within L2 contexts, Politeness and Impoliteness are best viewed in terms of appropriateness as not all L2 learners necessarily have sufficient linguistic command to intentionally manipulate language in order to convey such subtleties of meaning.

In the next sub-section, I detail the role of "hedges" in modifying the force of pragmatic meaning, in phenomena such as speech acts, conversational implicatures and politeness strategies.

1.2.3. Hedges

Hedges lie at the core of the semantic-pragmatic debate. Lakoff (1972) claims that natural language concepts have vague boundaries and fuzzy edges. As a result, "natural language sentences will very often be neither true, nor false, nor nonsensical, but rather true to a certain extent and false to a certain extent, true in certain respects and false in other respects" (Lakoff, 1972, p. 458). Lakoff defines "hedges" as words whose meaning implicitly involves fuzziness, i.e. "words whose job is to make things fuzzier or less fuzzy" (1972, p. 471).

Through his study of hedges, Lakoff draws very interesting conclusions with regards to the semantic- pragmatic interface. Some of his conclusions are as follows:

1. Hedges interact with felicity conditions and with rules of conversation as shown in the examples below (Lakoff, 1972, p. 490):

(1) "***Technically**, I said that Harry was a bastard*".

The above sentence is usually understood as "I said it but I didn't mean it". "*Technically*" appears to cancel the implicature that if you say something, you mean it.

(2) "***You might want to** close that window, Private Snurg*".

This situation conveyed in (2) may be understood as:

a) Strictly speaking, the sergeant did not order the private to close the window.
b) Essentially, the sergeant did order the private to close the window.

Thus, the hedging formula "*you might want to*" appears to weaken the degree of imposition of the command.

2. There may be hedges in lexical items such as in "*pink*", which may be viewed as a hedge between red and white.

3. The study of modifiers like "*sort of*", "*par excellence*", "*typical*", "*in essence*" and "*in a manner of speaking*" indicate that sentences asserting category of membership of an individual or an object, correspondingly display a degree of truth.

4. Semantics is not independent of Pragmatics. The study of the hedge "*regular*" (Bolinger, 1972) indicates that sentences with "*regular*" assert connotations. Traditionally, connotations are part of pragmatics and have nothing to do with truth conditions. Having said that, as the truth conditions of sentences with the hedge "*regular*" depend solely on connotations, "it follows that if connotations are part of pragmatics, then semantics is not independent of pragmatics" (Lakoff, 1972, p. 492). Furthermore, connotations are associated with the real-world situation and thus, are part of pragmatic information.

From a different perspective, Chafe (1986) addresses the notion of hedges within evidentiality. In a broad sense, "evidentiality" refers to linguistic expressions of attitude to

knowledge. Chafe (1986) indicates that people are usually aware, though not necessarily "consciously" aware, that some portions of their knowledge are truer or more accurate than others and that not all knowledge is equally reliable. Thus, such knowledge may be qualified with expressions that indicate speakers' assessment of its degree of reliability, the likelihood of its being a fact. Conversational English uses forms like adverbs such as "*maybe*", "*probably*" and "*certainly*" or modal verbs such as "*might*" and "*may*" for this purpose. Chafe (1986, pp. 266-269) draws a distinction between different modes of knowing: belief, induction, sensory evidence, hearsay evidence and deduction.

"Belief is a mode of knowing in which concern for evidence is downgraded" (Chafe, 1986). Belief forms include: "*I think*", "*I guess*" and "*I suppose*". "Induction" refers to a mode of knowing in which evidence plays a central role. The most common marker signalling an inference with a high degree of realiability is the modal verb "*must*". Other forms include the adjective "*obvious*", the verb "*seem*" and the adverb "*evidently*". Induction may be based on "sensory evidence". Markers of sensory evidence include explicit verbs like "*see*", "*hear*" and "*feel*" and phrases such as "*looks like*" and "*sounds like*". "Hearsay evidence" is a mode of knowing based on things people have told us about. Hearsay expressions include active and passive voice forms of reporting verbs like "*say*" and "*tell*" and the verbs "*suppose*" and "*seem*" and the adverb "*apparently*". "Deduction" is a mode of knowing in which one invents a model that predicts what will count as evidence. Markers of deduction include the modal verbs "*should*", "*can*", "*could*" and "*would*" and the adverb "*presumably*".

Furthermore, Chafe (1986, p. 270) argues that "human understanding involves a constant comparing of ongoing knowledge with expectations". In other words, we check everything we experience against things we already know, matching them with categories and schemas already present in our minds. Nonetheless, the matching of a piece of information with a category may not be that precise. In such cases, the knowledge has less than optimal codability. Markers of low codability include "*sort of*" and "*kind of*", which may qualify nouns, verbs, adjectives or whole predictions. Another hedge that removes the precision from categorisation but suggests something close is "*about*". Chafe (1986) also indicates that conversational English has a number of devices that signal expectations of some kind, against which knowledge may be matched. Such devices include "*of course*", "*oddly enough*", "*in fact*" and "*actually*".

Hedges are also commonly associated with politeness. Brown and Levinson (1987, p. 145) state that ordinary communication intentions, which are regulated and encoded in speech acts, are often potential threats to cooperative interaction. For instance, when one asks someone to do something, one presupposes that the latter is able and willing to do so, and has not done it yet. Thus, to hedge these assumptions is to avoid commitment to them. In other words, hegding "is a primary and fundamental method of disarming routine interactional acts" (Brown & Levinson, 1987, p. 146).

For Brown and Levinson (1987, p. 146), performative hedges are "the most important linguistic means of satisfying the speaker's want, don't assume H is able/willing to do A (and to some extent, the want to make minimal assumptions about H's wants)". Performative hedges might be analysed as adverbs on (often deleted) performative verbs that represent the illocutionary force of the sentence. Brown and Levinson (1987) also add that in some languages, performative hedges are encoded in words and particles that may also hedge propositional content. For instance, sentence (3) has two possible readings in the light of the adverb "*really*":

*(3) "He **really** did run that way".*

a) I tell you **sincerely**, he ran that way;
b) I tell you he **certainly** ran that way.

Brown and Levinson (1987, p. 162) indicate that there are numerous expressions in English which hedge illocutionary force, for example:

(4) "*That's just how it is, **in fact/in a way/in a sense/as it were/in all probability/I shouldn't be surprised/it seems to me/don't you agree***".

Furthermore, "if-clauses" enable felicity conditions to be suspended (Brown &Levinson, 1987, p. 162):

(5) "*Close the window **if you can/ if it closes/if it isn't closed yet/if you want***".

As for hedges addressed to Grice's Maxims, Brown and Levinson (1987, p. 164) state that quality hedges may suggest that the speaker is not taking full responsibility for the truth of his utterance, for instance, *"There is some evidence to the effect that"*. Conversely, quality hedges may strengthen S's commitment to the truth of his utterance or disclaim the

assumption to the point that S's assertion is to inform H, for example, "*With complete honesty I can say"* and *"As you know",* respectively (Brown & Levinson, 1987, p. 165). Quantity hedges give notice that not as much or not as precise information is provided as might be expected. Archetypal expressions include "*roughly*", "*more or less*" and "*approximately*" (Brown & Levinson, 1987, p. 166). In addition, there are expressions with clear politeness functions like "*I'll just say*", "*Well*", "*You know*" and "*I mean*".

In relation to Relevance hedges, Brown and Levinson (1987) argue that topic changes may be perceived as impositions on H's face, and thus, are done off record. Hedges marking the change in topic and partially apologising for it include forms like *"This may not be relevant/appropriate/timely, but", "Since I've been wondering"* and *"By the way"* (Brown & Levinson, 1987, p. 169). Lastly, Manner hedges include forms such as "*If you see what I am getting at*", *"To be succinct" and "In a nutshell"* (Brown & Levinson, 1987, p. 171). Manner hedges also query whether H is following S's discourse adequately[17] , for instance "*Yeah?"*, "*Got it*?" and "*You with me*?" (Brown & Levinson, 1987, p. 171).

Brown and Levinson (1987, p. 171) conclude that Maxim hedges are used with great frequency in ordinary talk and some have straightforward politeness applications. Firstly, Quality hedges that weaken S's commitment may redress advice or criticisms whereas those that strengthen are useful for making suggestions. Secondly, Quantity hedges can be used to redress complaints or requests. Thirdly, Relevance hedges may be used to redress offers and suggestions. Last but not least, Manner hedges can be used to redress all kinds of face threatening acts.

The previous views on hedges (Lakoff, 1972, Chafe, 1986, Brown & Levinson, 1987) highlight the difficulty in providing a fully comprehensive categorisation of hedging devices since their realisation include different forms such as adverbs, adverbial phrases, attitudinal verbs, modal verbs, conditionals and approximators, among others. In addition, such forms may be employed to perform different pragmatic functions. Nikula (1996) treats hedges as "pragmatic force modifiers". "Pragmatic force modifiers" encompass modifying devices which speakers use to either soften or strengthen the force of their messages (Nikula, 1996, p. 43). The author clarifies that the term "hedge" has been used either to refer to both softening

[17] Checking whether or not the listener is following ongoing discourse will be addressed in chapter 6 as self-imposed information receipt response tokens.

and strengthening devices (Brown & Levinson, 1987) or to softening devices (Prince et al., 1982) only. Thus, Nikula (1996) follows the position that hedges are a sub-type of pragmatic modifiers, used for mitigating and softening purposes.

Nikula (1996, pp. 50-55) draws a distinction between "explicit" and "implicit" pragmatic force modifiers based on the semantic-pragmatic divide. "Explicit modifiers" include attitudinal adverbs (e.g. "*presumably*", "*perhaps*"), parenthetical constructions (e.g. "*I think*", "*I suppose*") or adverbs of degree (e.g. "*a bit*", "*somewhat*", "*sort of*"). Such markers indicate quite explicitly the degree to which speakers are committed to the truth or preciseness of their messages. "Implicit modifiers" have a tendency to remain ambiguous or fuzzy even in the context, often making various interpretations possible. Pragmatic particles like "*I mean*", "*you know*", "*like*" and "*well*" are typical examples of implicit forms. Nikula (1996) clarifies that whilst both explicit and implicit modifiers are devices which speakers can use for interpersonal purposes, namely conveying feelings and attitudes and signalling politeness and involvement, they accomplish these purposes in different ways.

Explicit modifiers are transparent in that they convey quite clearly the speaker's attitude to the message. "This, in turn, usually has interpersonal functions such as the speaker's willingness to appear tactful and involved" (Nikula, 1996, p. 53). Implicit modifiers, on the other hand, are more directly related to the speaker's attitudes and relationship to the addressee. This relationship is often more opaque than that of explicit modifiers. Nikula (1996, p. 53) argues that the distinction between explicit and implicit modifiers should be seen "as a continuum along which certain choices are usually explicit (e.g attitudinal adverbs), whereas others tend towards the more implicit end of the continuum (e.g. pragmatic particles)". Moreover, pragmatic modifiers often occupy a position in between the continuum. It is also possible for a modifier to move along the explicit- implicit continuum in a way that in some contexts, it can function either as a softener or as an emphasiser whereas in other contexts it allows for different interpretations.

In the below conversation (6), two native speakers of English are talking about taxes. Speakers make an abundant use of pragmatic force modifiers (Nikula, 1992, p. 70):

(6)

S3: **I mean** it's a lot less [than the full rate three] hundred four hundred five hundred pounds **you know**
S1: [it's manageable yes erm] yeah a rebate of that kind is at least **kind of** manageable mm and if we think **I mean** the students take advantage of the community as much if no more like anybody else **you know** we're not **like** exempt [from the rest we just pay for it]
S2: [yeah **I suppose** so but I **still think** you could] once you finish studying **and whatever** you're still gonna **you know** you're gonna have to pay the full whack then and **I think** you're gonna make up for then I **just I think** they shouldn't make students pay **I mean I kind of agree** – I can see the point of a poll tax in general **and everything** even though I don't agree with it cos **I think** the better off you are you should pay more **and everything**.

Nikula (1996, pp. 51-52) explains that in the above conversation, the modifiers "*kind of*", "*I think*", "*I suppose*", "*and whatever*" and "*and everything*" are relatively transparent in signalling speakers' attitudes to their messages. In other words, such forms signal quite explicitly either speakers' certainty/uncertainty towards their message or their willingness to leave their messages vague. Conversely, the function of the modifiers "*I mean*" and "*you know*" seems to be more ambivalent as it is not closely tied to the literal meaning of these expressions. Thus, they are implicit forms in the sense that they leave room for different interpretations. Furthermore, the use of "*still*" by S2 can be argued to perform different functions. It may be used to emphasise the speaker's opinion or to diminish its degree of assertiveness (considering that the speaker is aware that other people have different opinions on the topic).

In this sub-section I have presented four views on hedges, which lie at the core of the semantic-pragmatic divide. For Lakoff (1972), hedges are words whose meaning implicitly involves fuzziness, i.e. they make things fuzzier or less fuzzy. For Chafe (1986), people are usually aware that some portions of their knowledge are truer or more accurate than others and that not all knowledge is equally reliable. Thus, speakers use hedging expressions to qualitfy such knowledge by indicating their assessment of its degree of reliability, the likelihood of its being a fact. For Brown and Levinson (1987), ordinary communication intentions are often potential threats to cooperative interaction. Therefore, speakers use hedges to disarm routine interactional acts. Hedges address Conversational Maxims and activate specific Politeness strategies. For Nikula (1996), hedges are a sub-type of pragmatic

modifiers used for mitigating and softening the force of their messages. While explicit hedges indicate quite explicitly the degree to which speakers are committed to the truth or preciseness of their messages, implicit modifiers have a tendency to remain ambiguous or fuzzy even in the context, often making various interpretations possible.

Now that I have described pragmatic phenomena directly related to the domain of the speaker, namely speech acts, conversational implicatures, politeness strategies and hedges, let us shift our focus to interactional acts performed by both participants within Conversation Analysis. For the purposes of this thesis, Conversation Analysis is addressed for its constructs but it is not used as a methodology.

1.3. CONVERSATION ANALYSIS

Conversation Analysis (CA henceforth) originated in the pioneering studies on structural organisation of everyday language use carried out by the sociologist Harvey Sacks in California, between 1964 and 1975 (Hutchby & Wooffitt, 2008) . His main collaborators were Emanuel Schegloff and Gail Jefferson. Sacks initiated a radical research programme whose purpose was to examine the levels of social order that could be revealed in the everyday practice of talking. The sociologist claimed that ordinary conversation may be a deeply ordered, structurally organised phenomenon and could be best investigated by the use of recorded data of naturally occurring talk.

Hutchby & Wooffitt (2008) clarify that Sacks's approach differentiates from the contemporaneous theory of speech acts (previously presented) in the sense that Searle's categorisation of illocutionary acts was decontextualised and his rules were grounded on intuition rather than in the observation of empirical examples of talk. Sacks, on the other hand, began with actual utterances with actual contexts. By analysing a corpus of telephone calls to a suicide prevention centre, Sacks drew three major conclusions (In: Hutchby & Wooffitt, 2008, pp. 17-19):

1. Utterances may be viewed as objects which speakers use to accomplish particular things in their interactions with others;
2. Talk can be seen as methodic;

3. Talk-in-interaction, which refers to the investigation of actual utterances in actual contexts, can be treated as an object of analysis in its own right, rather than simply as a window through which we can view other social processes or broader sociological variables.

In addition, Sacks also claimed that there is order at all points in talk-in-interaction. Thus, all the above insights serve as the methodological basis for CA (Hutchby & Wooffitt, 2008, p. 20):

1. Talk-in-interaction is systematically organised and deeply ordered;
2. The production of talk-in-interaction is methodic;
3. The analysis of talk-in-interaction should be based on naturally occurring data;
4. Analysis should not initially be constrained by prior theoretical assumptions.

According to Hutchby & Wooffitt (2008, p. 11), CA is the "systematic analysis of the talk produced in everyday situations of human interaction". Thus, "talk-in-interaction" lies at its core of investigation since it encompasses not only the analysis of everyday conversation but also a broader scope of interactional settings, which fall outside the common sense of conversation (Schegloff, 2007, pp. xiii-xiv). Research is based on transcribed audio recordings of naturally occurring interaction, i.e. activities that are situated in the natural unfolding of people's lives, as opposed to being prearranged, set up in laboratories or experimentally designed. Its main aims can be summarised as follows (Hutchby & Wooffitt, 2008, p. 12): "to discover how participants understand and respond to one another in their turns at talk, with a central focus on how sequences of actions are generated"; "to uncover the often tacit reasoning procedures and sociolinguistic competencies underlying the production and interpretation of talk in organised sequences of interaction".

Hutchby & Wooffitt (2008) indicate that CA is only marginally interested in language, as its real object of study is the interactional organisation of social activities. Words used in talk are not just studied as semantic units but rather as products or objects designed and used in terms of the activities being negotiated in the talk such as "requests", "complaints" and "suggestions". In addition, "the accomplishment of order, of sense, or coherence, in talk-in-interaction is seen as inextricably tied to the local circumstances in which utterances are produced" (Hutchby & Wooffitt, 2008, p. 12). Thus, CA also focuses on the production and

interpretation of talk-in-interaction as an orderly accomplishment that is oriented to by the participants themselves. In the next sub-sections, I will detail basic interactional phenomena comprising turn-taking organisation and sequence organisation.

1.3.1. Turn-taking

The organisation of turn-taking is one of the most fundamental organisations of practice for talk-in-interaction. For Schegloff (2007, p. 1), turn taking organisation works extremely effectively, producing long stretches of turns-at-talk which follow one another with minimised gap or overlap between them. However, turns do not follow one another in an identical shape and may be grouped in different ways. For Schegloff (2007, p. 1), a great deal of talk-in-interaction is better investigated with respect to action rather than topicality. Thus, analysts focus on what it is doing rather than what it is about. To illustrate this claim, Schegloff (2007, p. 2) presents the following extract:

(1.01) Virginia, 11:16-19
1 Mom: = hhh Whooh! It is so hot tuhnight. *Would somebody like some more ice tea.
2 ((*= voice fades throughout TCU))
3 (0.8)
4 Wes: Uh(b)- (0.4) I (‘ll) take some more ice.

The utterance “*Would somebody like some more ice tea*” is better understood as “making an offer” rather than being about the topic “ice tea” as the subsequent turn reveals an alternative acceptance to the offer.

Turns at talk can be seen as constructed out of units. “Turn-constructional units” (TCUs henceforth) are the building blocks out of which turns are fashioned (Schegloff, 2007, p. 3). They broadly correspond to linguistic categories like sentences, clauses, phrases, lexical items and can be identified by prosodic and intonational means. TCUs have two key features. Firstly, the property of “projectability”, which refers to the possibility for participants to project, in the course of a TCU, what sort of unit it is and at what point it is likely to end (Hutchby & Wooffitt, 2008, p. 50). This feature leads to a second key property of TCUs known as “transition-relevance place”.

"A speaker beginning to talk in a turn has the right and obligation to produce one TCU, which may realise one or more actions" (Schegloff, 2007, p. 4). When a speaker approaches the possible completion of a first TCU, the transition to a next speaker can become relevant then be accomplished just after the completion of the TCU in progress. This span is referred to as "transition-relevance place". Schegloff clarifies that when a speaker approaches the TCU completion, this speaker transition does not necessarily occur but becomes possibly relevant:

> "Speakers often produce turns composed of more than one TCU…if a speaker talks past a possible completion of the first TCU in a turn, whether by extending that TCU past its possible completion or by starting another TCU, whether in the face of beginning of talk by another or clear of such overlapping talk, then at the next occurrence of imminent possible TCU completion transition to a next speaker again becomes possibly relevant" (Schegloff, 2007, p. 4).

Sacks et al. (1974) propose rules to describe how turns come to be allocated at transitional-relevant places. At the initial transition-relevance place of a turn (In: Hutchby & Wooffitt, 2008, p. 51):

Rule 1: (a) If the current speaker has identified, or selected, a particular next speaker, then that speaker should take a turn at that place;

(b) If no such selection has been made, then any next speaker may (but need not) self-select at that point. If self-selection occurs, then the first speaker has the right to the turn;

(c) If no next speaker has been selected, then alternatively the current speaker may, but need not, continue talking with another TCU, unless another speaker has self-selected, in which case that speaker gains the right to the turn.

Rule 2: Whichever option has operated, then rules 1 (a-c) come into play again for the next turn-transition place.

For Hutchby & Wooffitt (2008), these rules are intended as descriptions of the practices that participants display an orientation to in actual, local occasions of turn-taking. These rules are learned and tacitly known by speakers and therefore instantiated and reproduced in talk-in-interaction. However, how these rules operate in the temporal unfolding of talk-in-interaction is a matter for the participants. In addition, research on the apparent violations of this rule-set shows that those apparent violations are actually robust illustrations of how closely members do orient to the rules such as in overlapping talk and repair sequences.

Levinson (1983) highlights that around five per cent of the speech stream is delivered in "overlap", which describes moments in which two speakers speak simultaneously. Although overlapping talk may be considered evidence of incoming speakers' failure to take notice of whether prior speakers have or have not finished their turn, conversational analysts have demonstrated that most instances of overlap occur in the environment of possible transition-relevance places (Hutchby & Wooffitt, 2008). Jefferson (1986) argues that the onset[18] and the termination of overlap are extremely ordered and that the occurrence of overlap demonstrates speakers' effort to orient themselves to the rules of turn-taking. She identified three major categories of overlap onset (In: Hutchby & Wooffitt, 2008, p. 56):

1. **Transitional onset:** when a next speaker orients to a possible transition-relevance place;
2. **Recognitional onset:** when the next speaker feels they recognise what the current speaker is saying and can project its completion, even if that is before the end of a TCU;
3. **Progressional onset:** when there is some disfluency in the current turn and a next speaker suggests a completion in order to move the conversation forward.

"Repair" is a generic term used in CA to cover a number of phenomena, from seeming errors in turn-taking such as those involved in overlap , to any of the forms of correction, i.e. substantive faults in the contents of what a speaker has just said (Hutchby & Wooffitt, 2008, p. 57). The first sense of "repair" is used due to the fact that one way of seeing what is going on is in terms of repair of the turn taking system. The second sense is an alternative term to "correction" since not all conversational repair involves any factual error on the speaker's part. Thus, in this sense, "repair" involves the suspension of ongoing turns or sequences in order to attend to some trouble that has arisen.

Sacks et al. (1974) demonstrated two ways in which repair illustrates participants' orientations to the basic turn-taking rules. Firstly, the turn-taking system itself incorporates its own means of repairing faults. In other words, when there is overlap, there is a violation of the ideal of "one speaker at a time". However, this is repaired when one speaker stops speaking before the completion of a first TCU. Secondly, another form of repair is a procedural type in which

[18] In phonology, a syllable can be divided into three parts: the onset, its beginning; the nucleus, its central part; the coda, its end (Richards et al., 1990, p. 283).

participants orient to various aspects of the ongoing talk in order to manage turn-taking problems. Sacks et al. (1974) claim that participants display their continuing orientation to the turn-taking set in their management of these problems.

Schegloff, Jefferson and Sacks (1977) describe four varieties of repair sequences:

1. Self-initiated self- repair: both initiated and carried out by the speaker of the trouble source;
2. Other-initiated self-repair: carried out by the speaker of the trouble source but initiated by the recipient;
3. Self-initiated other-repair: the speaker of the trouble source may try and get the recipient to repair the trouble;
4. Other-initiated other-repair: the recipient of the trouble-source turn both initiates and carries out the repair; this is closest to what is conventionally understood as "correction".

Hutchby & Wooffitt (2008) stress that repair is closely connected to the temporal organisation of talk-in-interaction. There are a number of positions in which a repair initiation may occur in relation to the trouble source: within or immediately after the TCU containing the trouble source; immediately after the next transition-relevance place after the trouble source; in the turn following the turn containing the trouble source, i.e. in the next speaker's turn. However, the authors argue that all of the positions in which repair usually occurs are in close proximity to the trouble source.

Another interactional phenomenon associated with turn-taking is back-channelling. Yngve (1970) indicates that when speakers are engaged in conversations, they usually take turns. Following this, both participants in the conversation, i.e. the person holding the turn and his partner are simultaneously engaged in both speaking and listening. "This is because of the existence of what I call the backchannel, over which the person who has the turn receives short messages such as *yes* and *uh-huh* without relinquishing the turn" (Yngve, 1970, p. 568). In other words, listening partners use backchannels to indicate that they are following what is being said and also to inform how the message is being received. Backchannels invoke different communicative choices for listeners, enabling them to offer relevant feedback to speakers (Mott & Petrie, 1995). Speakers then engage in subsequent actions based on such

feedback. Backchannels can occur in both verbal (e.g. minimal and non- minimal forms) and non-verbal modes (e.g. gestures, laughter and gaze) providing speakers with different types of feedback. Verbal backchannels within a linguistic domain will be addressed in chapter 6 as a means for expressing "good listenership" (McCarthy, 2002).

Now that I have described phenomena related to turn-taking, namely TCUs, transition-relevance place, turn-taking rules, overlapping talk, repair and backchannels, let me address a central organising format of sequences called "the adjacency pair".

1.3.2. Pair-adjacency

According to Schegloff (2007), one very large set of sequence types appears to be organised around a basic unit of sequence construction called the "adjacency pair". The author clarifies that there are sequence organisations not based on adjacency pairs such as story-telling and some forms of topic talk. However, a larger number of sequences in talk-in-interaction are produced by reference to the practices of adjacency pair organisation. Therefore, adjacency pair organisation serves as a resource for sequence construction comparable to the way TCUs serve as a resource for turn construction.

"Minimal, basic unexpanded adjacency pairs" display the following features (Schegloff and Sacks 1973, pp. 295-296):

1. Composed of two turns
2. by different speakers
3. adjacently placed; that is, one after the other
4. These two turns are relatively ordered; i.e. they are differentiated into "first pair parts" ("Fs" henceforth) and "second pair parts" ("Ss" henceforth). "Fs" are utterance types such as "question", "request" and "invitation" as these initiate some exchange. "Ss" are utterance types such as "answer", "grant", "reject", "accept" and "decline" as these are responsive to the action of the prior turn.
5. The components of an adjacency pair are pair-type related, i.e. not every second pair part can properly follow any first pair part. Adjacency pairs compose pair types such as "greeting-greeting", "question-answer" and "offer-accept/decline".

The basic rule of operation by which the minimal form of the adjacency pair is produced works as follows (Schegloff, 2007, p. 14): given the recognisable production of a first pair part, on its first possible completion its speaker should stop, a next speaker should start and produce a second pair part of the same pair type:
A: "*How do you do*". (First pair part- greeting)
B: "*How do you do*". (Second pair part- greeting)

Schegloff (2007) indicates that an adjacency pair in its basic, minimal two-turn form can itself constitute the whole of a sequence and is common, virtually formulaic, in the opening and closing sections of conversations and other types of episodes of talk-in-interaction. Nevertheless, Schegloff (2007) argues that none of the above features of adjacency pairs (and the rule of operation) is rigid or invariant. Adjacency pair-sequences can have more than two turns, though they will still possess the two basic parts; they can be separated by intervening talk known as "insert expansions"; some utterance types may be used as both first pair parts (Fs) and second pair parts (Ss) , for instance, a complaint might be used to initiate a sequence but also in response to an inquiry; and under specified circumstances, some utterance types may be used as both Fs and Ss at the same time, e.g. when a person asks the speaker to repeat a question, by repeating the question, the speaker grants the request (second pair part) and asks the question again (first pair part).

The expansion of minimal adjacency pairs occurs in the three possible places that a turn unit allows: before the first pair part as a "pre-expansion", between the first pair part and the second pair part as an "insert expansion" and after the second pair part as a "post-expansion". Firstly, "pre-expansions" are themselves constructed of adjacency pairs and are regularly referred to as "pre-sequences" (Sacks, 1992). They are themselves sequences and come before sequences, being preliminary to something else. This "something" else they are preliminary to is quite specific and is the first pair part of a particular pair type such as an invitation or an offer. Preliminaries that project such imminent first pair parts are called "type-specific pre-sequences" such as "pre-invitations", "pre-offers", "pre-requests" and "pre-announcements". Schegloff (2007) adds that pre-sequences seem largely designed to avoid various kinds of troubles in the sequence which a prospective first pair part would start. Let us examine a sequence provided by Schegloff (2007, p. 30):

(4.01) JG 3:1 (Nelson is the caller; Clara is called to the phone)

1 Cla: Hello
2 Nel: Hi.
3 Cla: Hi.
4 Nel: F (pre) Whatcha doin'.
5 Cla: S (pre) Not much.
6 Nel: F (b) Y'wanna drink?
7 Cla: S (b) Yeah.
8 Nel: Okay.

In line 4, Nelson's utterance works as a pre-invitation for a drink. Clara's response in line 5 encourages Nelson to go ahead with the invitation with the base first part pair, which the "pre" was projecting. Nelson then makes the invitation in line 6, which is followed by Clara's acceptance in line 7. Acceptance to an invitation is a "preferred response". Other possible responses to the pre-invitation include a blocking response, such as "*I am actually studying for a test*", which would discourage the invitation and avoid a refusal , or a hedging response, such as "*Well, we are going out. Why*", which would make the full response dependent on what the invitation was. Thus, pre-sequences are used to anticipate and to avoid potential troubles such as a refusal, a rejection or a disagreement, or in other words, "dispreferred responses".

The design of preference organisation was originally accounted for by Pomerantz (1984). Pomerantz (1984, p. 64) states that isolated turn-and-sequence shapes enable different kinds of actualisations of the actions being performed with and through them. She characterises two types of shapes: "preferred-action turn shape" and "dispreferred-action turn shape". "Preferred-action turn shape" is designed to maximise the occurrences of the actions being performed with them. It minimises the gap between its initiation and prior turn's completion and comprises components that are explicitly stated instances of the action being performed. "Dispreferred-action turn shape" , on the other hand, "minimises the occurrences of the actions being performed with them, in part utilising the organisation of delays and nonexplicitly stated action components, such as actions conditionally relevant next" (Pomerantz, 1984, p. 64).

Pomerantz (1984) argues that there is an association between an action's preference status and the turn shape in which it occurs. Schegloff (2007, p. 60) explains that interactional projects and course of actions are implemented in sequence organisation in such a way that:
a) Plus (+) responses are "preferred" such as acceptances, grantings and agreements;
b) Minus (-) responses are "dispreferred" such as rejections, declinings and disagreements.
Nevertheless, Schegloff (2007) clarifies that there are many exceptions to this general preference structure. For instance, agreement is not always the preferred response in the case of self-deprecation and pro-forma offers. Under such circumstances, disagreement would be the preferred response.

In addition, Schegloff (2007) argues that some response types (second pair parts) to some first pair parts do not seem to fit into the plus/minus categorisation. For instance, agreeing or disagreeing with an announcement suggests a stance toward the accuracy of what has been told, i.e. the propositional content of the utterance. However, the generic issues for announcement sequences relate to the recipient's alignment towards the first pair part as news (in opposition to "already know") and towards the assessment of the news as good or bad. Thus, the alignment with the designed presentation of the news by the recipient of an announcement (the announcement informs the recipient of news, which can be good or bad) can be considered "preferred agreement" and non-alignment as "dispreferred rejection".

Returning our discussion to the second type of expansion of minimal adjacency pairs, "insert expansions" are also constructed out of adjacency pairs. As for their placement, they occur after the base first pair part and before the base second pair part, as illustrated in the example below:

(6.01) SBL 2,1,8 (Schegloff et al., 1977, p. 368)
1 Bet: Fb- Was last night the first time you met Missiz Kelly?
2 (1.0)
3 Mar: Fins- Met whom?
4 Bet: Sins- Missiz Kelly.
5 Mar: Sb- Yes.

The previous example highlights two indispensable elements for the characterisation of insert expansions: position between a first pair part and a projected second pair part; the insert expansion is initiated by the recipient of the preceding first pair part. Schegloff (2007) adds that the initiation of an insert sequence is aimed at delaying the base second pair part. Insert sequences are delivered so as to address matters that need to be dealt with in order to enable the doing of the base second pair part. In other words, insert sequences project the doing of the base second pair part upon completion of some preliminary work. They carry the base second pair part forward and renew it. Once an insert sequence is finalised, the base second pair part becomes relevant again. In addition, as insert sequences intervene between first and second base pair parts, by compromising the progressivity of the base sequence, they project the possibility of a dispreferred response.

Lastly, some sequences are recognisably complete (to speakers and to observers) at the end of the second pair part turn. This is particularly noticeable in sequences that have no preference structure such as greetings and leave-takings or bye-byes. However, sequences can be expanded after the delivering of the second pair part as "post expansions". Schegloff (2007, p. 117) explains that sequences with preferred or agreeing or + [plus] second pair parts are "closure-relevant" whilst ones with dispreferred or disagreeing or – [minus] second pair parts are "expansion-relevant". In other words, preferred responses usually lead to closing the sequence whereas dispreferred ones regularly lead to the expansion of the sequence.

In this sub-section I have investigated preference structure, the three different positions minimal adjacency pairs can be expanded and the practices involved in such expansions. For Schegloff (2007), preferred or dispreferred responses refer to a structural relationship of sequence parts rather than participants' motives, desires or likings. Pre-sequences are built out of minimal adjacency pairs and are commonly used to anticipate and to avoid dispreferred responses. Insert sequences are also constructed by minimal adjacency pairs and are produced so as to address matters that need to be dealt with in order to enable the doing of the base second pair part. Post inserts refer to the expansion of sequences after the delivering of the base second pair part and may be used to guide a sequence to its closing and/ or to license the start of a next sequence.

\

SUMMARY

In this chapter, I have presented the foundations of Discourse Analysis, Pragmatics and Conversation Analysis and detailed discourse, pragmatic and interactional phenomena inherent in face-to-face conversations. For the purposes of this thesis, I chose to adopt a functional approach to discourse analysis (Brown & Yule, 1983), which focuses on the analysis of language in use and describes linguistic forms along with their real life functions. Language was acknowledged to perform both transactional and interactional functions, namely the communication of factual and propositional information and the establishment and maintenance of social relationships, respectively.

The production of spoken language is more challenging to language producers when compared to written language as speakers need to monitor and adjust ongoing discourse, manage turns and repair sequences in real time. However, speakers are able to rely on interlocutors' immediate reactions to modify their discourse (Brown & Yule, 1983). The grammar of spoken conversation is linear since it proceeds incrementally, and dynamic as it takes place in real time (Culpeper & Kytö, 2010). It includes specific multi-word combinations or lexical bundles and specific structural patterns such as prefaces (discourse markers, overtures, fronting), bodies (extending) and tails (tag questions, comment clauses, retrospective hedges and noun phrase tags).

As for the organisation of discourse, cohesion refers to the relations of meaning that exist within a text and that define it as a text (Halliday and Hasan, 1976). Such cohesive relationships may be achieved by means of reference, substitution, ellipsis, conjunctions and lexical cohesion. A text is a passage of discourse that is coherent in two dimensions: with respect to the context in the situation (being consistent in register) and with respect to itself (being cohesive). Continuatives are a marginal category of items that do not fit into the proposed taxonomy of conjunctions but may be used with a cohesive force in the text. On describing continuatives and their cohesive uses, Halliday and Hasan (1976) seemed to be talking about discourse markers.

Discourse markers or pragmatic markers display the following features (Brinton 1996, pp. 33-35): a high frequency feature of oral discourse, sometimes with more than one occurrence

in a single sentence; short items, often phonologically reduced or unstressed; restricted to sentence-initial position, though they may appear sentence-medially and finally; having little or no propositional meaning; occurring either outside the syntactic structure or loosely attached to it; optional rather than obligatory features. Discourse markers are a subtype of pragmatic markers (Fraser, 1996) and have the additional feature of mediating discourse (Culpeper & Kytö, 2010). They link segments of discourse in ways which reflect choices of monitoring, organisation and management by speakers and mark speakers' stance or attitude towards the message (Carter & McCarthy, 2006).

The second part of this chapter was devoted to Pragmatics and pragmatic phenomena. Pragmatics as meaning in interaction (Thomas, 1995) was the perspective adopted since making meaning is a dynamic process, involving the negotiation of meaning between the speaker and the hearer, the context of an utterance and its meaning potential. Speech acts are utterances produced by speakers with the intention of performing actions through language (Austin, 1962). Illocutionary acts can be classified into five mutually exclusive and jointly exhaustive classes (Searle, 1976): representative/assertive, directive, commissive, expressive and declarative. The notion of indirect speech acts (Searle, 1975, 1979) was initially characterised as cases in which one illocutionary act is performed indirectly by way of performing another and then revisited so as to encompass Grice's conversational implicatures (1975).

Grice's (1975) cooperative principle was detailed. The flouting of conversational maxims (Quantity, Quality, Relation and Manner) indicates that a speaker is trying to say something else beyond the conventional meaning of the sentence uttered. Conversational implicatures are inferences that arise to preserve the assumption of co-operation. Generalised conversational implicatures arise without any particular context or special scenario whereas particularised conversational implicatures require specific contexts (Grice, 1975). Utterance-type-meaning is a level of systematic pragmatic inference based on general expectations about how language is normally used rather than on direct computations about speaker-intentions (Levinson, 2000). It was claimed that speech acts, felicity conditions, conversational pre-sequences and generalised conversational implicatures operate at this intermediate level of meaning (Levinson, 2000).

Politeness (Brown & Levinson, 1978, 1987) is a universal principle and a pre-condition of human cooperation. It encompasses universal features like face (negative and positive) and rationality. Face threatening acts are acts produced by speakers that intrinsically threaten face, namely those acts that by their nature run contrary to the face wants of the addressee (H) and/ or of the speaker (S). When speakers perform FTAs, they rely on four super-strategies: bald on record (direct forms that may include mitigating devices or hedges), positive politeness (strategies oriented towards the positive face of H), negative politeness (respecting H's freedom of action) and off record (the flouting of particularised conversational implicatures).

Hedges lie at the core of the semantic-pragmatic debate and perform different functions. Definitions of hedges include: words whose meaning implicitly involves fuzziness (Lakoff, 1972); expressions used by speakers to qualify some portions of their knowledge as more or less accurate or true (Chafe, 1986); devices to disarm routine interactional acts (Brown & Levinson, 1987). For the purposes of this thesis, I narrowed the notion of hedges to a sub-type of pragmatic modifiers (Nikula, 1996) used for mitigating and softening the force of speakers' messages. Explicit hedges indicate quite explicitly the degree to which speakers are committed to the truth or preciseness of their messages while implicit modifiers have a tendency to remain ambiguous or fuzzy even in the context, often making various interpretations possible (Nikula, 1996).

The third and final part of this chapter addressed Conversation Analysis and interactional practices. Conversation Analysis is a sociological approach that views ordinary conversation as a deeply ordered, structurally organised phenomenon and utterances as objects that speakers use to accomplish particular things in their interactions with others (Hutchby & Wooffitt, 2008). Participants orient themselves to turn-taking rules and apparent violations like overlapping talk and repair sequences are actually robust illustrations of how closely members orient to such rules (Sacks et al., 1974, Jefferson, 1986). Backchannels are verbal (and non-verbal) devices used by listening partners to indicate that they are following what is being said and also inform how the message is being received (Yngve, 1970).

Adjacency pair organisation serves as a resource for sequence construction (Schegloff, 2007). Minimal adjacency pairs are composed of two pair-type related turns, adjacently placed by different speakers (Schegloff & Sacks, 1973). They are virtually formulaic in the opening and

closing sections of conversations (Schegloff, 2007). The expansion of minimal adjacency pairs occurs in the three possible places that a turn unit allows: before the first pair part as a pre-expansion, between the first pair part and the second pair part as an insert expansion and after the second pair part as a post-expansion. Pre-sequences are employed to anticipate and to avoid dispreferred responses. Preferred responses usually lead to closing the sequence whereas dispreferred ones regularly lead to the expansion of the sequence.

In the next chapter, I will endeavour to relate the above phenomena and practices to interlanguage pragmatics and competence frameworks within L2 contexts. By doing so, I aim at highlighing key discourse, pragmatic and interactional phenomena for a conversational competence model in L2 within a pragmatic perspective. Furthermore, as discourse markers, hedges and backchannels perform pragmatic functions, it seems appropriate to regard them as sub-types of pragmatic markers. Thus, the functions of some specific discourse markers, hedges and backchannels will be investigated with learner and native speaker data in chapters 4, 5 and 6 respectively.

2. INTERLANGUAGE PRAGMATICS AND COMPETENCE FRAMEWORKS

In section 1.2, we discussed the scope of pragmatics and presented pragmatic theories based on two major pragmatic notions. Firstly, when speakers produce utterances, they also perform acts through them (Austin, 1962, Searle, 1969). Secondly, speakers rely on a cooperative principle governed by conversational Maxims in order to implicate beyond what is said (Grice, 1975, Brown & Levinson, 1987). Due to their philosophical nature, the above theories focus mainly on L1 contexts. Röver (2005) suggests that it was Leech (1983) who provided the most convenient starting point for relating pragmatics to second language acquisition (SLA henceforth) by narrowing the scope of his discussion to general pragmatics and by describing its culturally bound facets: pragmalinguistics and socio-pragmatics. Pragmalinguistics can be applied to "the study of the more linguistic end of pragmatics-where we consider the particular resources which a given language provides for conveying particular illocutions" (Leech, 1983, p. 11). Socio-pragmatics, on the other hand, "is the sociological interface of pragmatics" (Leech, 1983, p. 10). Leech (1983) argued that the Cooperative Principle and the Politeness Principle clearly operate variably in different cultures or language communities, in different social situations and among different social classes.

Following from Leech (1983) and Thomas's (1983) two-fold division of pragmatics, pragmalinguistic competence encompasses interactants' knowledge of strategies for realising speech intentions and the linguistic items used to express these intentions (Röver, 2005). In other words, pragmalinguistics refers to the resources speakers use for conveying communicative acts and relational or interpersonal meanings (Kasper, 1997). Sociopragmatic competence comprises knowledge of the social conditions governing language use (Röver, 2005). It presupposes participants' awareness of degrees of relative power, social distance and the degree of imposition involved in a specific communicative act (Brown & Levinson, 1987) and of their rights and obligations (Thomas, 1983). According to Röver (2005), it is rather difficult to draw a clear line between what belongs to each domain when analysing performance data. Pragmalinguistic and sociopragmatic competences are intrinsically intertwined for two main reasons. Firstly, language use is invariably contextual. Secondly, both competences are involved in producing and in comprehending speech intentions. However, Röver (2005) suggests that this division is theoretically and empirically useful and

that research shows that learners can be more advanced in one of these competences than in the other.

Selinker's notion of interlanguage (1972) formalises the interplay between SLA and pragmatics. Gass and Selinker (2008) indicate that SLA research assumes that learners create a language system known as an interlanguage. This system has its own structure and is composed of numerous elements, not the least of which are elements from the native language and the target language. In addition, there are elements that do not originate from either the native language or the target language. The latter elements are called new forms and are the empirical essence of interlanguage. Gass and Selinker (2008, p. 14) state that "what is important is that the learners themselves impose structure on the available linguistic data and formulate an internalised system". Conversely, the concept of fossilisation refers to the cessation of learning. Due to the difficulty in establishing when learning has ceased, some SLA researchers favour the term stabilisation of linguistic forms. Gass and Selinker (2008) conclude that in SLA, it is noticeable that some learners reach plateaus that are far from the target language norms. Furthermore, it seems that fossilised or stabilised interlanguages exist despite learners' further exposure to the target language.

In this chapter, I focus on the second language learner by addressing Interlanguage Pragmatics and competence frameworks that are embedded in the interface between SLA and Pragmatics. The objectives of this chapter are as follows:

1. To revisit the construct "conversational competence" in L2 and place it within a pragmatic domain, grounded on the notion of "meaning in interaction" (Thomas, 1995);
2. To detail key discourse, pragmatic and interactional phenomena which comprise conversational competence in L2.

In order to pursue the above objectives, I initially present views on the domains of Interlanguage Pragmatics (Kasper & Blum-Kulka, 1993, Kasper & Rose, 2002, Schauer, 2009) and address areas of interlanguage investigations. I also summarise two SLA theories (Schmidt, 1993, Bialystok, 1993) that account for pragmatic learning. I then focus on SLA more specifically by critically revisiting communicative competence frameworks (Canale & Swain, 1980, Canale, 1983, Bachman, 1990) and definitions of interactional competence

(Schmidt, 1983) and conversational competence (Richards & Sukwiwat, 1985, Thornbury and Slade, 2006, Corsetti, 2009). Next, I justify and propose a conversational competence model in L2 within a pragmatic domain, highlighting key discourse, pragmatic and interactional phenomena previously presented in chapter 1. Lastly, I relate the proposed model of conversational competence in L2 to the level descriptors of the Common European Framework for Language Reference (Council of Europe, 2001) and justify theoretical and methodological choices.

2.1. INTERLANGUAGE PRAGMATICS

Kasper and Blum-Kulka (1993, p. 3) indicate that "Interlanguage Pragmatics" (ILP henceforth) is a "second-generation hybrid" as it belongs to two different interdisciplinary fields. As a branch of Second Language Acquisition Research, ILP is one of its specialised sub-areas, contrasting with interlanguage studies in phonology, morphology, syntax and semantics. ILP within SLA domain examines how L2 learners develop the ability to understand and perform action in a target language (Kasper & Rose, 2002, p. 5). As a sub-set of Pragmatics, it can be regarded as a socio-linguistic, psycholinguistic or simply linguistic enterprise, depending on the scope of pragmatics adopted. ILP within a pragmatic domain investigates how non-native speakers comprehend and produce action in a target language (Kasper & Rose, 2002, p. 5).

ILP is traditionally regarded as the study of non-native speakers' use and acquisition of linguistic action patterns in a second language (Kasper & Blum-Kulka, 1993, p. 3). However, narrowing the scope of ILP to non-native speakers or language learners can be restrictive. Based on the study of American immigrants in Israel, Blum-Kulka (1991) demonstrates that "speakers fully competent in two languages may create an intercultural style of speaking that is both related to and distinct from the styles prevalent in the two substrata, a style on which they rely regardless of the language being used" (Kasper and Blum-Kulka, 1993, p. 3). Kasper and Blum-Kulka (1993) clarify that the intercultural style hypothesis is supported by many studies of cross-cultural communications, in particular within interactional sociolinguistics and research into the pragmatics of immigrant populations across generations. While most ILP studies investigate non-native speakers, reflecting the status of ILP as a branch of second

language research, Blum-Kulka and Sheffer (1993) extend the perspective to include native speakers' intercultural styles.

Schauer (2009, p. 15) places ILP as a relatively young area in linguistics, which originated from pragmatic theories and developments in L2 pedagogy and research in the 70's. Schauer (2009) indicates that ILP uses pragmatic theories, principles and frameworks to examine how L2 learners encode and decode meaning in their L2. The author adds that ILP research is also heavily influenced by Hymes's (1972) concept of communicative competence, which resulted in a development away from a more grammar-centred L2 pedagogy. His model of communicative competence and the subsequent ones by Canale and Swain (1980), Canale (1983) and Bachman (1990) contributed to a shift in L2 pedagogy towards communicative language learning and teaching, which in turn, resulted in an increased interest in pragmatic studies focusing on language learners in the mid 1980s. Communicative competence frameworks will be addressed in the next sub-section.

Let us now focus on the different domains of ILP investigation, namely pragmatic comprehension, production of linguistic action, pragmatic development, pragmatic transfer and pragmatic failure. Early studies on pragmatic comprehension focused on learners' attribution of illocutionary force and perception of politeness. Kasper and Blum-Kulka (1993) point out that research on the attribution of illocutionary force has centred on the comprehension of indirect speech acts, factors contributing to facilitate or strain the understanding of pragmatic comprehension, the role of linguistic form and context information and learner variables influencing force attribution. Politeness inquiries have investigated how learners assess the politeness value of different speech act realisation strategies through off-line metapragmatic judgement tasks such as card sorting, paired comparison, multiple-choice and rating scales. Kasper and Blum-Kulka (1993) indicate that the results of such studies confirm that learners are capable of distinguishing different degrees of politeness in conventions of means and forms. However, their perceptions do not always match those of native speakers. Some studies have also focused on non-native speakers' socio-pragmatic perceptions such as universal and culture-specific assumptions about apology frequency and realisation.

As for pragmatic production, evidence suggests that "regardless of a particular L1 and L2, and of the type of learning context (naturalistic vs. instructed), learners have access to the same range of realisation strategies for linguistic action as native speakers, and demonstrate sensitivity to contextual constraints in their strategy choice" (Kasper & Blum-Kulka, 1993, p. 7). Nonetheless, learners face difficulties in exploiting such general pragmatic knowledge base, due to their restricted L2 linguistic knowledge, or in accessing it smoothly. Moreover, a lack of L2 pragmalinguistic sophistication, combined with negative transfer of socio-pragmatic norms from L1 (i.e. the influence of L1 on interlanguage pragmatic knowledge that differs from the L2 target) or non-native perceptions of L2 socio-pragmatic norms, or even purposeful loyalty to L1 cultural norms, may result in deviations from native use even at high proficiency levels (Blum-Kulka, 1991).

Research on learners' distribution patterns of strategies and forms also indicates variability from those of native speakers. While some studies report on learners' preference for more direct modes of expressing pragmatic meaning than those of native speakers, others suggest the opposite. The quality and range of linguistic forms by which linguistic action can be implemented and modified has consistently been demonstrated to differ between non-native speakers and native speakers. Non-native speakers' repertoire is typically more restricted and less complex than that of native speakers (e.g. Scarcella, 1979, Blum-Kulka, 1991, Schmidt, 1983). Some of these findings may simply reflect learners' lexical and syntactic knowledge whereas others clearly indicate a pragmalinguistic issue, in situations when learners know a particular lexical item or syntactic structure but use it in a way that does not convey the intended illocutionary force or politeness value.

According to Kasper and Blum-Kulka (1993), most ILP research focuses on pragmatic comprehension and production, rather than on pragmatic development. The investigation of pragmatic development requires more longitudinal studies in naturalistic settings, including close observation of learners from the very beginning of their language acquisition process. The authors also highlight the importance of a theoretical framework to account for pragmatic learning. Within cognitive approaches to ILP development, Schmidt (1993) and Bialystok (1993) are highly influential theoreticians.

Schmidt's noticing hypothesis (1993, p. 27) assumes that linguistic forms can serve as intake for language learning provided such forms are noticed by learners. While paying attention to such forms is helpful, there are other factors in the learning context which focus attention on them so that they are noticed. In addition, general principles of the organisation of language may be discovered through the use of explicit or implicit learning mechanisms. Even in situations "where what native speakers know about the pragmatic principles of their language is inaccessible to consciousness, such knowledge may nevertheless be based on insights and understanding at the time of learning" (Schmidt, 1993, p. 27).

Schmidt (1993, pp. 35-36) makes very important claims with regards to the learning process in itself and the learning of pragmatics in a second language as opposed to Krashen's (1981) view that implicit learning is superior to conscious problem solving:

a) Attention to input is a necessary condition for any learning itself;
b) What must be attended to is not input in general but whatever features of the input play a role in the system to be learned;
c) For the learning of pragmatics in a second language, attention to linguistic forms, functional meanings, and the relevant contextual features is required;
d) Learners experience their learning;
e) Attention is subjectively experienced as noticing;
f) The attention threshold for noticing is the same as the threshold for learning;
g) While incidental and implicit learning are both possible, consciously paying attention to the relevant features of input and attemping to analyse their significance in terms of deeper generalisations are both highly facilitative;
h) There is a stronger relationship between motivation, acculturation and other affective factors in the development of pragmatic and discoursal ability than in other aspects of language learning, such as sytax;
i) Explicit teacher-provided information about the pragmatics of a second language can also play a role in learning, on condition that it is accurate and not based solely on fallible native speaker intuitions.

In a classroom-based project with Brazilian learners of English, Corsetti (2009) investigated the role of strategy-based listening (Mendelsohn, 1995, 1998) in the enhancement of pragmatic comprehension and production. Her findings corroborated the following

assumptions: in order to achieve listening proficiency, learners need practice in making inferences as semantic and pragmatic inferences are embedded in verbal communication; semantic and pragmatic aspects affecting the meaning of utterances can be highlighted via comprehension activities focusing on specific listening sub-skills. The listening comprehension activities learners were exposed to included the activation of pragmatic knowledge embedded in the dialogue exchanges. In addition, subjects' overall level of listening proficiency improved considerably during the five-month classroom project, which may be accounted for by the fact that learners were given a number of opportunities to infer pragmatic meanings, among other listening micro-skills. However, there was no significant improvement in learner's pragmatic production. Corsetti's results (2009) are aligned with Schmidt's position (1993) that close attention to the relevant features of input facilitates pragmatic learning. Her findings highlight the importance of paying attention to contextual clues and functional meanings, particularly in the development of pragmatic comprehension.

In another cognitive view, Bialystok (1993, p. 43) indicates that pragmatic competence comprises speakers' ability to use language for different purposes (e.g. to request, to instruct, to effect change) and listeners' ability to get past the language and understand speakers' real intentions. It also includes a command of the rules by which utterances are strung together to create discourse. In order to produce coherent speech, speakers rely on turn-taking, cooperation and cohesion, among other factors. Bialystok (1993, p. 49) divides the elements that are required for the analysis of linguistic systems into three levels of representation: conceptual, formal and symbolic. "In conceptual representation, language is organised only around the meanings it represents" (Bialystok, 1993, p. 49). Thus, utterances can be formed to satisfy specific speech acts but speakers are focusing on the intended meanings and not on the forms being selected to express such intentions. "Formal representations" are coded in terms of the structure of the language and the categories that organise the representation are the formal categories of language itself. At this level, speakers are capable of classifying a group of utterances as being requests or assertions and are aware of the relation between direct and indirect speech acts. "In symbolic representation, the coding is a relation between a form and a referent"(Bialystok, 1993, p. 49). There is additionally an explicit representation for the relation between the pragmatic intention and the forms used to achieve that intention. Thus, at this level, learners would have direct access to a range of alternative forms of expression which indicate more or less the same intentional content.

Following from this, non-native speakers need to learn some social conventions for the new language and to solve some mapping problems between forms and social conditions. However, "the task of organising a linguistic system along the lines of speech functions so that the system is operable in contexts has been solved for a first language (Bialystok, 1993, p. 54). Thus, adjustments to this symbolic system, which carries pragmatic functions, is a relatively small problem. The main challenge for adult learners is to develop the control strategies to attend to the intended interpretations in contexts and to select the forms from a range of possibilities, which satisfy the social and contextual needs of the communicative situation. "Pragmatic competence is achieved when control of processing is mastered for a richly analysed representation of the language" (Bialystok, 1993, p. 54).

Perna (2002) analysed Brazilian learners of English as a second language producing apologies and showed that they had difficulties producing the appropriate form according to the degree of formality. In her study, she shows that learners tend to overgeneralise the formula "*I´m sorry*" in detriment to more formal ones, required for the specific social context. She suggests that strategies for developing this communicative competence should be incorporated in the L2 learning process.

Now that we have addressed pragmatic comprehension, production and development, let us focus on the remaining domains of ILP investigations, namely pragmatic transfer and pragmatic failure. The notion of pragmatic transfer is closely related to both cognitive theories previously addressed (Schmidt, 1993, Bialystok, 1993). Kasper and Blum-Kulka (1993) highlight that learners' native language and culture influence their interlanguage pragmatic knowledge and performance. Most studies tend to focus on negative transfer as it potentially risks communicative success. Negative transfer refers to the influence of L1 pragmatic competence on interlanguage pragmatic knowledge that differs from the L2 target. Positive transfer, on the other hand, refers to pragmatic behaviours or other knowledge displays which are consistent across L1, interlanguage and L2. Positive transfer usually results in communicative success. Kasper and Blum-Kulka (1993) clarify that at a conceptual level, negative transfer does not necessarily reflect lack of competence in the pragmatics of the target community.

When non-native speakers communicate in a style that differs from that of native speakers, it is important to consider if this style is seen as lacking in some way, or just different, and if its maintenance over time is considered negatively, as fossilisation, or positively, as a marker of cultural identity. “The degree of socio-cultural accommodation to the L2 culture may be as well a matter of choice as of ability” (Kasper & Blum-Kulka, 1993, p. 11). A foreign accent, for instance, may shield non-native speakers as it identifies them as such and thus, may lower native speakers’ expectations vis-à-vis non-native speaker linguistic behaviour. In addition, the desirable goal for highly proficient L2 speakers may be that of disidentification, rather than of absolute convergence, in contexts of immigration or in the use of L2 in cross-cultural communication.

Furthermore, grammatical or phonological interlanguage deviations from the target language are easily recognised by native speakers and thus, are not usually attributed to flaws in non-native speakers’ personality (Kasper & Blum-Kulka, 1993). Pragmatic failure, on the other hand, may not necessarily be perceived as faulty learning by interlocutors and may result in miscommunication. Blum-Kulka and Olshtain (1986) view pragmatic failure as linked to cultural variability in the implementation of Grice’s (1975) conversational model: certain types of pragmatic deviations from the target norms, such as learners’ tendency to verbosity, are seen as violations of a cultural norm for the balance required between the maxims of clarity and quantity. Thomas (1983, p. 91) defines pragmatic failure as “the inability to understand what is meant by what is said”. Pragmatic failure results in communication breakdowns among interlocutors of different cultural backgrounds. Some reasons why utterances may be incorrectly inferred are as follows (Thomas, 1983, p. 94):

a) Hearers perceive the force of speakers’ utterances as stronger or weaker than that of the intended ones;
b) Hearers perceive speakers’ utterances as ambivalent where speakers intended no ambivalence;
c) Speakers expect hearers to be able to infer the force of their utterances relying on a system of knowledge or beliefs that is not shared by speakers and hearers.

Kasper and Blum-Kulka (1993) add that although communication strategies are clearly recognisable as pragmatic areas, they are usually included under ILP. While the study of communication strategies has predominantely been grounded in psycholinguistic models of

cognitive processing, ILP has focused on general and cross-cultural pragmatics. Research on communication strategies has investigated learners' solutions to referential problems whereas ILP has examined the illocutionary and politeness dimensions of speech act performance. Kasper and Blum-Kulka (1993) conclude that while this division of labour reflects different research traditions, it has little theoretical support as attested by Bachman's (1990) framework of communicative competence.

To conclude, we have seen that ILP is a "second-generation hybrid" belonging to SLA and Pragmatics (Kasper & Blum-Kulka, 1993). Within SLA domains, ILP investigates how L2 learners develop the ability to understand and perform actions in a target language (Kasper & Rose, 2002, p. 5). Within a pragmatic perspective, ILP examines how non-native speakers comprehend and produce action in a target language (Kasper & Rose, 2002, p. 5). ILP specifically focuses on non-native pragmatic comprehension, the production of linguistic action, pragmatic development, negative and positive transfer and pragmatic failure. In the next sub-section, I present a critical review of communicative competence frameworks (Canale & Swain, 1980, Canale, 1983, Bachman, 1990) which have greatly influenced ILP studies.

2.2. COMMUNICATIVE COMPETENCE FRAMEWORKS

According to Dell Hymes (1972, p. 281), a person who acquires communicative competence acquires knowledge and ability for language use with respect to whether (and to what degree) something is: formally possible; feasible in virtue of the means of implementation available; appropriate (adequate, happy, successful) in relation to a context in which it is used and evaluated; and is in fact done, actually performed, and what its doing entails. Hymes designed a programme for the integrated study of language and use, i.e. the ethnography of speaking. For Hymes (1972, p. 53), to describe communicative competence, one needs to consider the speech community, the speech situation, the speech event, the speech act, the fluency of the speaker, the components of speech events and the functions of speech.

Communicative competence constructs in L2 usually include at least a code and a use component, of which pragmatic and discourse competences are sub-categories. The code component describes a language user's knowledge of syntax, morphology, semantics, lexis,

and phonology whereas the use component relates to the ability to use language appropriately for a purpose within a given context (Niezgoda & Röver, 2001). Among the most influential frameworks for the characterisation of communicative competence are Canale and Swain's (1980) and Canale's (1983) and Bachman's (1990), who adapted Dell Hymes's first language model (1972) to a second language construct on testing.

Canale and Swain (1980) initially sub-divided "communicative competence" into three sub-competences:

1. Linguistic or grammatical competence: consists of the knowledge of the basic elements of communication such as sentence patterns, morphological inflections, lexis and phonological or orthographic systems;
2. Sociolinguistic competence: consists of the social and cultural knowledge required to use language appropriately with reference to formality, politeness and other contextually defined choices; it refers to the degree sentences are produced and understood appropriately;
3. Strategic competence: includes the strategies and procedures relevant to language learning, language processing and language production. It activates knowledge of the other competences and helps language users overcome communication difficulties.

The above model was refined by Canale (1983), who introduced a fourth component:

4. Discourse competence: refers to the knowledge of how to combine grammatical forms and meanings to achieve a unified spoken or written text in different genres and situations.

The discourse proponents Celce-Murcia and Olshtain (2000, p. 16) regard discourse competence as the central component since "this is where everything else comes together". In other words, the realisation of all of the other competences is done through discourse. Scarcella and Oxford (1992) accept Canale and Swain's framework of communicative competence with two major revisions. Firstly, Scarcella and Oxford (1992, p. 72) extend the notion of discourse competence to refer to "verbal, nonverbal, and paralinguistic knowledge underlying the ability to organise spoken and written texts meaningfully and appropriately". The authors highlight that some researchers favour the term "conversational competence" to

refer to this broader definition of discourse competence related to conversations. Secondly, Scarcella and Oxford (1992) expand the original concept of "strategic competence" to include all types of compensation strategies that make up for missing knowledge such as guessing from context in reading and listening and paraphrasing and circumlocution[19] in speaking and writing.

According to Csépes (2009), the above four-part model has two major problems. Firstly, it fails to clearly specify how communicative competence is translated into communicative performance. Secondly, it does not indicate how the different components interact with one another, apart from suggesting that strategic competence operates in relation to the other competences. McNamara (2000, p. 18) argues that strategic competence is oddly named as it is not a type of stored knowledge but "a capacity for strategic behaviour in performance, which is likely to involve non-cognitive issues such as confidence, preparedness to take risks, and so on". McNamara (2000) also states that discourse competence has elements of a general intellectual flexibility in negotiating meaning in discourse, rather than being only stored knowledge.

Bachman (1990) proposes a "theoretical framework of communicative language ability", which attempts to shed light on the above issues. For Bachman (1990, p. 84), the ability to use language communicatively involves "both knowledge of, or competence, and the capacity for implementing it, or executing that competence in appropriate, contextualised communicative language use". Therefore, his framework of communicative language ability includes three components: language competence, strategic competence and psychophysiological mechanisms (Bachman, 1990, p. 85).

[19] Circumlocution refers to using too many words to express oneself, usually to avoid saying something clearly.

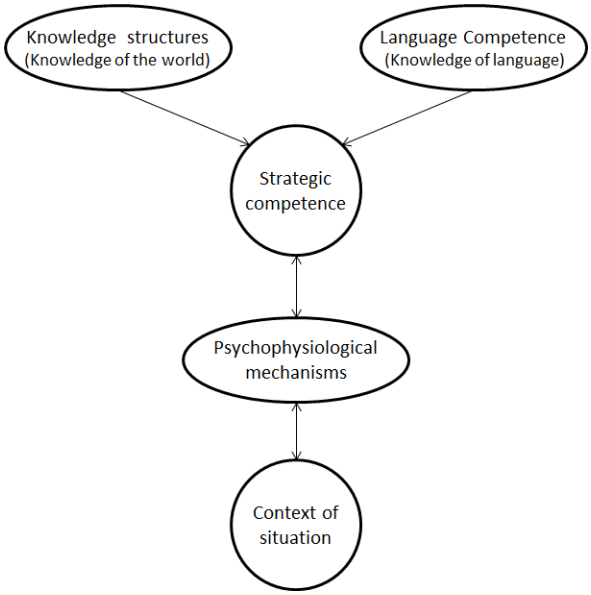


Figure 1: Framework of communicative language ability (Bachman, 1990)

Communicative competence is seen as "a dynamic system in which world knowledge and language competence feed into strategic competence, which defines the degree to which linguistic intentions are efficiently executed" (Niezgoda & Röver, 2001, p. 64). Strategic competence then interacts with "psychophysiological mechanisms", which refer to "the neurological and psychological processes involved in the actual execution of language as a physical phenomenon" (Bachman, 1990, p. 84). Lastly, these mechanisms interact with the context of a particular situation.

The first component "language competence" comprises two major competences: "organisational competence" and "pragmatic competence". Both language sub-competences consist of several categories which interact with one another and also with features of the language use situation (Bachman, 1990, p. 87):

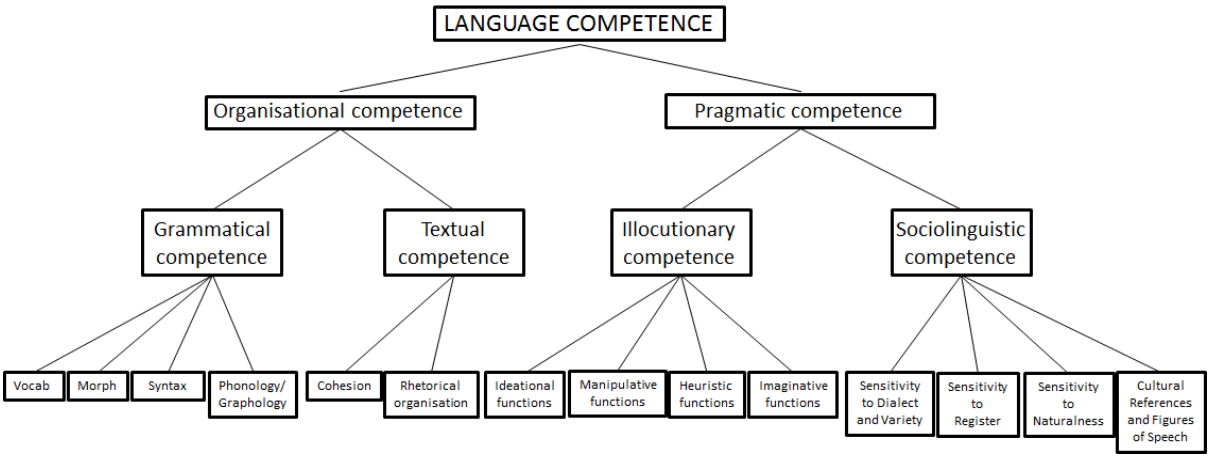


Figure 2: Language competence and its sub-competences (Bachman, 1990)

"Organisational competence" encompasses grammatical and textual abilities. "Grammatical competence" consists of a number of independent competences that are involved in language usage[20] such as the knowledge of vocabulary, morphology, syntax, and phonology/graphology. "Textual competence" includes the knowledge of rules governing text formation. Bachman (1990) adds that a text is essentially a unit of language (spoken or written) consisting of two or more utterances or sentences that are structured together according to rules of cohesion and rhetorical organisation. Therefore, "textual competence" also comprises conversational language use and conventions involved in establishing, maintaining and terminating conversations.

"Pragmatic competence" or "pragmatic knowledge", as revised by Bachman and Palmer (1996, p. 69), "enables us to create or interpret discourse by relating utterances or sentences and texts to their meanings, to the intentions of language users, and to relevant characteristics of the language use setting". In Bachman's original framework (1990), "pragmatic competence" encompasses two features: illocutionary competence and sociolinguistic competence. "Illocutionary competence" aids the interpretation of the relationships between utterances or sentences and texts and the intention of language users (Bachman & Palmer, 1996). According to Kasper (1997), it refers to the knowledge of the communicative action (in both written and spoken modes) and how to carry it out. "Illocutionary competence" or "functional knowledge", as revised by Bachman and Palmer (1996, pp. 69-70), comprises knowledge of four categories of language functions:

[20] "Language usage" refers to the function of a linguistic item as an element in a linguistic system" (Widdowson, 1978).

1. Ideational: functions that enable us to express or interpret meaning in terms of our experience of the real world and include the use of language to express or exchange information about ideas, knowledge or feelings. Utterances performing these functions include descriptions, classifications, explanations, and expressions of anger and sorrow;

2. Manipulative: they enable us to affect the world around us. They include instrumental functions (getting other people to do things, for example, "requests" and "suggestions"), regulatory functions (controlling what other people do, for instance, "rules" and "regulations") and interpersonal functions (establishing, maintaining and changing interpersonal relationships like "greetings" and "compliments");

3. Heuristics: functions that enable us to use language to extend our knowledge of the world around us, for instance, when using language for teaching and learning, for solving problems, and so on;

4. Imaginative: they enable us to use language to create an imaginary world or extend the world around us for humorous or aesthetic purposes such as jokes, figures of speech and poetry.

Bachman and Palmer (1996) conclude that these four categories of language functions are by no means mutually exclusive. They do not usually occur only in individual or isolated utterances. On the contrary, most language use involves the performance of multiple functions in connected discourse. The following quote summarises Bachman's original views on these functions and how they relate to sociolinguistic competence:

> While illocutionary competence enables us to use language to express a wide range of functions, and to interpret the illocutionary force of utterances or discourse, the appropriateness of these functions and how they are performed varies from one language use context to the next, according to a myriad of sociocultural and discoursal features. (Bachman, 1990, p.94).

"Sociolinguistic competence" is "the sensibility to, or control of the conventions of language use that are determined by the features of the specific language use context" (Bachman, 1990, p. 94). Therefore, it fosters the performance of language functions in ways that are appropriate to that context. It encompasses the following abilities: sensitivity to differences in dialect or variety, sensitivity to differences in register, sensitivity to naturalness and ability to interpret cultural references and figures of speech.

All things considered, Bachman's (1990) framework places the notion of pragmatic competence as a sub-category of language competence as the former comprises the ability to use language appropriately according to a specific situational context. Bachman claims that attempts to validate the various components of different communicative competence constructs have not been conclusive enough while Bachman and Palmer (1982) were able to demonstrate how grammatical and pragmatic competences are closely associated with each other.

The second component "strategic competence" consists of "a set of metacognitive components, or strategies, which can be thought of higher order executive processes that provide a cognitive management function in language use, as well as in other cognitive activities" (Bachman & Palmer, 1996, p. 70). For the authors, using language involves the language user's topical knowledge and affective schemata along with all the areas of language knowledge previously described. "Topical knowledge" refers to knowledge structures stored in long-term memory while affective schemata relate to their affective or emotional correlates. The integration of all these components makes language use possible as language users create and interpret discourse in situationally appropriate ways.

Lastly, "psychophysiological mechanisms" refer to the neurological and psychological processes that operate at the execution phase of language use (Bachman, 1990). They characterise the channel (auditory or visual) and the mode (receptive or productive) in which competence is implemented. In receptive language use, auditory and visual skills are activated whereas in productive language use, neuromuscular skills are activated. In their revised model, Bachman and Palmer (1996) do not adopt "psychophysiological mechanisms" as a third component of communicative language ability but address them under the topic of "language skills". For the authors, language skills may not be part of the language ability itself but rather the contextualised realisation of the ability to use language in the performance of specific language use tasks.

For Csépes (2009), Bachman's original framework (1990) represents a major reconceptualisation as it locates "competence" within a wider "performance" framework and it separates strategic competence from language competence. For McNamara (1997), Bachman's discussion of communicative language ability (1990) gives a central place to the

notion of "interaction" in language performance. However, McNamara (1997) stresses that Bachman uses the term "interaction" to refer exclusively to cognitive activity on the part of the candidate, which implies viewing the latter in a curious kind of isolation. For McNamara (1997), the successful performance in language tests also depends on other factors such as interlocutors, elicitation methods and the interpretation of performance. In his view, Bachman (1990) and Bachman and Palmer (1996) fail to account for the co-constructed nature of communicative performance.

To sum up, we have investigated two communicative competence frameworks in this section. Canale and Swain (1980) and Canale's (1983) model was the first attempt to adapt Dell Hymes's first language framework for second language testing. It consists of four independent competences (linguistic, sociolinguistic, strategic and discourse competences) and it does not indicate how communicative competence is translated into communicative performance or how the different components interact with one another either. Bachman's dynamic framework (1990) is based on the principle that communicative language ability involves both knowledge of language and the capacity for implementing it. It comprises three components (language competence, strategic competence and psychophysiological mechanisms) which interact with one another and with the context of a particular situation. However, it does not seem to consider the co-constructed nature of communicative performance.

2.3 CONVERSATIONAL COMPETENCE

In the previous characterisations of communicative competence, the notion of "conversational competence" is placed within a broader framework, which accounts for both oral and written discourse. Canale's (1983) fourth component "discourse competence" relates to the knowledge of how to combine grammatical forms and meanings to achieve a unified spoken or written text in different genres and situations. Similarly, Bachman (1990) places conversational language use under "textual competence", a sub-competence of "organisational competence", as a text refers to both written and oral modes of discourse. These theoretical positions raise an important question. Does the ability to produce spoken discourse require similar efforts by language producers as those required in written texts?

As previously detailed in section 1.1.1., oral and written discourse differ tremendously vis-à-vis the demands they make on language producers (Brown & Yule, 1983). Sinclair and Coulthard's (1975) classes of conversational moves and the prototypical characteristics of spoken conversation (Culpeper & Kÿto, 2010) highlight the uniqueness of spoken discourse, especially when one considers its real-time processing and interactivity. As Halliday (1985, p. xxiii) states, "writing exists, whereas speech happens". Thornbury and Slade (2006, p. 25) define "conversation" as "the informal, interactive talk between two or more people, which happens in real time, is spontaneous, has a largely interpersonal function, and in which participants share symmetrical rights". Thus, the ability to produce spoken discourse and interact with other speakers seems complex enough to justify a separate characterisation from the ability to produce written discourse, especially if we consider the negotiation of meaning which takes place in real time.

The negotiation of meaning through spoken discourse is accounted for in the communicative competence frameworks previously described. In Bachman's model (1990), pragmatic competence mediates the production and the interpretation of discourse by relating utterances or sentences and texts to their meanings, to the intentions of language users, and to relevant characteristics of the language use setting. By subdividing pragmatic competence into sociolinguistic and illocutionary sub-competences, Bachman also echoes Leech (1983) and Thomas's (1983) sociopragmatic and pragmalinguistic dimensions respectively. Canale and Swain (1980) and Canale (1983) also consider the appropriateness of utterances in their description of sociolinguistic competence. However, they also indicate that strategic competence comprises verbal and non-verbal strategies such as paraphrasing and asking for clarification, which are called into action to compensate for breakdowns in communication.

By assuming the need for a separate categorisation for spoken discourse that includes the negotiation of meaning, a second question arises. What does the ability to produce spoken discourse in interactional settings encompass? To avoid drawing premature conclusions, let us initially examine some previous definitions of interactional competence and conversational competence as both terms may overlap at times.

2.3.1. Definitions of interactional and conversational competence

According to Kasper and Rose (2002), the notion of "interactional competence" emerged simultaneously in different social sciences. Heritage and Atkinson (1984) indicate that the main aim of Conversation Analysis is to describe and explicate the different competences speakers use and rely on when engaged in intelligible, socially organised interaction. Gumperz (1982, p. 209), a proponent of interactional sociolinguistics[21], redefines "communicative competence" in interactional terms as "the knowledge of linguistic and related communicative conventions that speakers have to sustain and create conversational cooperation". Gumperz (1982, p. 209) lists conditions by which "interactional competence" may be acquired: the need for face-to-face contact in situations that allow for maximal feedback; the existence of outside conditions such as ethnic and class solidarity, which assist the learning of discourse by forcing interlocutors to disregard breakdowns and stay in contact or give the learner the benefit of the doubt.

In a second language acquisition study, Schmidt (1983) describes the acquisition of English without formal instruction by an adult Japanese learner called Wes during a three-year period in Hawaii. He analyses the learner's accomplishments based on Canale and Swain (1980) and Canale's (1983) four-part model of communicative competence. When detailing discourse competence, Schmidt focuses on spoken discourse as Wes does not write in English. Schmidt indicates that spoken discourse other than monologue is a cooperative effort by all parties in a conversation and, thus, addresses both interactional competence and conversational competence as sub-parts of discourse competence. Schmidt illustrates Wes's "interactional competence" in terms of extended routines when ordering food at a restaurant, the coherence and the use of speech markers in narratives and descriptions and the use of rhetorical features. Schmidt also exemplifies the learner's "conversational competence" in terms of sequence organisation, conversational style and turn-taking routines. Schmidt stresses that "discourse competence" was the area in which Wes had the most significant improvements over time. By the end of the observational period, Wes was able to carry out small talk appropriately and to use language –specific backchannels in English and in Japanese.

[21] Interactional sociolinguistics is a linguistic sub-discipline that investigates how language users create meaning via social interaction (Gumperz, 1982).

For Richards and Sukwiwat (1985), "conversational competence" describes the speaker's knowledge of how speech acts (Austin, 1962, Searle, 1969) are used in social situations and is defined with reference to utterances. They also indicate that the investigation of how utterances are used shows that many have a recurring and predictable nature and are associated with particular social situations and with particular types of interactions. These utterances may be referred to as "conversational routines" and may include several different types of conventional utterances. Richards and Sukwiwat (1985) argue that many social events and speech acts require the use of conversational routines. These help define speech situations and their appropriate use is a vital component of social competence in a language.

Conversational routines have been investigated in a number of studies such as Coulmas (1981), Schauer & Adolphs (2006) and Paquot & Granger (2012). Nonetheless, it is Aijmer (1996) who provides the most substantial treatment of conversational routines by looking at them in speech and relating them to pragmatic phenomena. Conversational routines or discoursal expressions include a range of frequent phrases in spoken language such as swear words, exclamations, greetings, polite responses, discourse-organising formulae and small talk (Aijmer, 1996) . They can be grouped into different classes such as formulaic speech acts (direct and indirect speech acts, simple speech acts and routinised patterns of speech acts), discourse organising functions (connectives and conversational gambits), attitudinal routines (routines that express speakers' attitudes or emotions) and idioms. In her study, Aijmer (1996) adopts a functional approach to the study of conversational routines and describes different types of routinised speech acts in the 435,000-word London-Lund Corpus of Spoken English. She also investigates speech acts with the function of thanking and apologising, routinised requests and offers and routines that have a discourse-strategic function.

Aijmer (1996, p. 11) defines "conversational routines" as "phrases, which, as a result of recurrence, have become specialised or 'entrenched' for a discourse function which predominates over or replaces the literal referential meaning". Aijmer (1996, p. 11) argues that conversational routines from different semantic sources can be functionally equivalent as exemplified below:

Let's face it
To put it mildly
To tell you the truth

Frankly
Honestly
Forgive me if I say that…
I don't want to be personal but…

The above phrases occur at the beginning of an utterance or larger discourse unit and they can be pragmatically grouped as different ways to warn the hearer that the forthcoming message may be experienced as indiscreet, embarrassing or blunt.

Aijmer (1996) explains that conversational routines can minimally consist of a single word, though most routines comprise several words. Some routines are discontinuous requiring a lexical insertion, for instance, *as far as/…/ is concerned*. Routines do not only occur alone and they can be repeated or juxtaposed. The emerging structures are composed of single building blocks that are stored as units in the speaker's long-term memory, appearing in the order in which they are generated. It is also possible to find sequences of routines. The smallest of the formulaic discourse structures can be described in terms of pair-adjacency[22].

Greetings (repeated routines/adjacent pair):
A: How do you do.
B: How do you do.

Expressing gratitude (juxtaposed routines):
A: Lovely. Thank you very much indeed!

Thanking (minimal adjacent pair)
A: Thank you very much.
B: Don't mention it.

As for indirect speech acts, Aijmer (1996) regards indirect requests (e.g. "*Can you pass me the salt"*) as routine forms just as much as "*thank you"* and "*sorry"*, among others. In an indirect speech request, there is a mismatch between form and function. However, the speaker can use semi-lexicalised units such as "*Can you…".* Aijmer (1996) clarifies that some

[22] Pair adjacency was addressed in section 1.3.2.

indirect requests may be pragmatically ambiguous and, thus, the routine "*Can you…"* can be either a request formula or a question formula. Another feature of indirect speech acts is that they can be more or less explicit. The use of a performative formula maximises the explicitness of the illocutionary force. Conversely, the use of an explicit or transparent marker, which is only indirectly associated with a certain function, weakens such explicitness. Following this perspective, "*I'm sorry"* is an apology marker just as much as "*I apologise"*, although the former is less explicit.

In Aijmer's study (1996), the rules for the use of indirect speech acts are culturally-specific. She clarifies that while the strategies are universal or general, there might be cross- linguistic differences between indirect speech acts. Aijmer (1996) suggests that the differences found among speech acts are, to a large extent, motivated by different cultural norms rather than by the strategy being used. The formula "*Can't you…"*, for instance, is regulated by language specific social rules. Leech (1983, p. 123) analysed the sense of "*Can't you shut up"* and spelt it out semantically as follows: "I have assumed that you cannot shut up. I now doubt whether this is true. I ask you to resolve my doubt". Aijmer (1996) explains that the above utterance can be intended as an ironic reproach. Conversely, the formula "*Can't you…*" in Danish is less restricted and may be used as a polite request routine. The same applies to Brazilian Portuguese in which the negative form is largely used to introduce polite requests.

Now that we have described conversational routines, let us present two further views on conversational competence. Thornbury and Slade (2006, p. 188) define "conversational competence" as "the subset of linguistic and strategic competencies that are implicated in conversation (as opposed to writing, or other non-conversational spoken registers)" according to their working definition of conversation (informal, spontaneous, interactive talk between two or more people, happening in real time, with a largely interpersonal function and in which participants share symmetrical rights). Conversational competence draws on both textual knowledge, including the knowledge of conversational organisation, and on pragmatic knowledge, including both functional and sociolinguistic knowledge.

In addition, Thornbury and Slade (2006, p. 214) indicate that "the way conversational skills develop in an L2 is not the same as the way they develop in the L1". Unlike children who are learning their mother tongue, second language adult learners already have a well-developed

conversational competence in their mother tongue, which can be applied to L2 contexts. Good conversationalists know how to open and close conversations, take turns, change topic, interrupt politely, hold the floor, fill pauses, indicate discourse direction via signpost words, use backchannels, tell stories and be cooperative in their mother tongue. Therefore, the authors go on to describe the factors that facilitate or hinder the transfer of those sub-skills into a second language and the relation between conversational competence and learners' overall language proficiency.

In order to specifically investigate what gets in the way of second language learners understanding pragmatic meaning, Corsetti (2009) proposes an alternative framework for the characterisation of pragmatic competence in verbal communication. In this alternative model, "pragmatic competence" encompasses three components: inferential competence (representing the notion of pragmatic comprehension), conversational-interactional competence (representing the notion of pragmatic production) and sociolinguistic competence (representing the notion of appropriateness and interacting with both comprehension and production dimensions).

Corsetti's (2009) inferential competence construct is based on the assumption that pragmatic comprehension involves the comprehension of speech acts and implicatures (Thomas, 1995) and, more broadly speaking, the interpretation of pragmatic inferences. Corsetti (2009) defines "inferential competence" as the ability to interpret pragmatic meaning such as the comprehension of speech acts (Austin, 1962, Searle, 1969) and implicatures (Grice, 1975, Levinson, 2000), and is characterised by a taxonomy of illocutionary acts, the assignment of illocutionary force, conventional implicatures and generalised and particularised conversational implicatures.

With regards to pragmatic production, Corsetti (2009) defines "conversational-interactional competence" as speakers´ ability to produce illocutionary acts conveying the intended illocutionary force as well as the ability to manage dialogic and monologic discourse[23]. Managing dialogic discourse involves opening and closing conventions; making small talk; turn-taking conventions (recognition of transition relevance place); using adjacency-pairs; holding the floor (hesitation fillers and floor holding devices); interrupting politely (using

[23] The skills involved in managing dialogic and monologic discourse derive from Richards (2006).

functional exponents and overlapping); reacting to others (paralinguistic features such backchannels and prosody). Likewise, managing monologic discourse includes using an appropriate register (oral presentations, lectures, speeches); presenting information in an appropriate sequence (using signpost words); maintaining audience engagement (coherence of ideas).

As for appropriateness, Corsetti (2009) adopts Canale and Swain (1980) and Bachman's (1990) notions of "sociolinguistic competence". Sociolinguistic knowledge affects both conversational-interactional and inferential sub-competences and acts as a filter for successful communication. While speakers need to select appropriate forms and be aware of polite linguistic behavior in L2 in order to convey their intended meanings successfully, listeners also need to be aware of socio and cultural references and polite routines in L2 in order to be able to interpret pragmatic meaning.

At the time of her study (2009), Corsetti's main motivation was to characterise what got in the way of learners comprehending and producing pragmatic meaning. As her proposal included an empirical investigation, it seemed theoretically valid to separate pragmatic production from pragmatic comprehension so as to examine which area posed more difficulties for her learners, who were about to take the IELTS examination and were admittedly struggling with the listening component. The initial results of a classroom project confirmed that pragmatic comprehension was the most challenging dimension of pragmatics for her learners and, thus, a methodological approach (strategy-based listening) was subsequently implemented. Nonetheless, bearing in mind a more comprehensive treatment of conversation, it seems essential for a "conversational competence" construct to encompass both production and comprehensions dimensions and, as a result, include the notion of appropriateness.

In this section I have presented different views on conversational competence. In Canale and Swain (1980) and Canale's (1983) and Bachman's (1990) models, conversational language use is placed within broader frameworks, which account for both oral and written discourse, namely discourse competence and textual competence, respectively. Based on Canale's discourse competence (1983), Schmidt (1983) describes interactional competence in terms of extended routines, the coherence and the use of speech markers and the use of rhetorical features whereas he defines conversational competence in terms of sequence organisation,

conversational style and turn-taking routines. For Richards and Sukwiwat (1985), conversational competence describes the speaker's knowledge of how speech acts are used in social situations and is defined with reference to utterances or conversational routines. For Thornbury and Slade (2006, p. 188), conversational competence is "the subset of linguistic and strategic competencies that are implicated in conversation" and draws on both textual and pragmatic knowledge. In Corsetti's (2009) three-part framework for pragmatic competence, conversational-interactional competence represents the notion of pragmatic production and refers to speakers´ ability to produce illocutionary acts conveying the intended illocutionary force as well as the ability to manage dialogic and monologic discourse. In short, the above definitions of conversational competence include elements derived from Discourse, Pragmatics and Conversation Analysis studies. However, they do not seem to specify the interplay between discourse, pragmatic and interactional phenomena or to account for both comprehension and production dimensions.

In the next sub-section, I introduce a revised view of conversational competence and detail key discourse, pragmatic and interactional phenomena which seem to be intertwined during conversations. I abandon the adjective "interactional" so as to avoid overlap with sociological and sociolinguistic terminology but I account for the interactional aspects of conversation within the pragmatic domain.

2.3.2. Conversational competence within a pragmatic domain

As previously mentioned in chapter 1, the study of conversation is traditionally associated with Discourse Analysis, Pragmatics and Conversation Analysis. However, each field of study focuses on different aspects of conversation. The reason for placing conversational competence within a pragmatic domain is grounded on two assumptions: conversation is prototypical of language usage (Levinson, 1983) and language in use involves the negotiation of meaning. Thomas (1985) indicates that the negotiation of meaning involves the participation of both speakers and hearers and the consideration of the context of utterances and their meaning potential. Thus, speakers make use of linguistic and paralinguistic means to convey their intended meanings, to repair breakdowns in communication, to establish and maintain social relationships, to negotiate the exchange of turns, peer solidarity and face saving acts, among other conversational phenomena. Listeners also play an active role in the

sense that they signal whether they are following the ongoing discourse or require clarification. They encourage speakers to continue talking through response tokens. Lack of response is also meaningful as it expresses lack of interest in continuing the ongoing conversation.

Furthermore, as previously mentioned in section 1.1, language is used for both transactional and interactional purposes (Brown & Yule, 1983). Thus, speakers use language to convey factual or propositional information and to establish and maintain social relationships, respectively. In Bachman and Palmer's (1996) framework of communicative competence, second language learners use language to perform ideational functions (expressing or exchanging information about ideas, knowledge or feelings), manipulative functions (getting other people to do things, controlling what other people do and establishing, maintaining and changing interpersonal relationships) , heuristics (extending their knowledge of the world around them) and imaginative functions (creating an imaginary world or extending the world around them for humorous or aesthetic purposes), assisted by illocutionary or functional knowledge. Bachman and Palmer (1996) conclude that the above four categories of language functions are by no means mutually exclusive and that most language use involves the performance of multiple functions in connected discourse. Their model of communicative competence acknowledges the role of illocutionary knowledge in the performance of language functions.

From my empirical observation as an EFL teacher, when learners engage in oral interactions, some of them do particularly better than others at communicating their intended messages. Performance varies from person to person, according to their overall linguistic ability, to their mastery of the different aspects of conversational competence, to individual personal characteristics, to their relationship with interlocutors, among other factors. Obviously, outgoing individuals are more likely to feel at ease during conversations provided they feel linguistically prepared for tasks and/or are familiar with the topics addressed. Shier individuals face the further challenge of overcoming affective barriers. In addition, when listening partners do not show much interest in what speakers are saying, the latter may feel discouraged from moving beyond the initial stages of their conversation. Nevertheless, there are occasions on which learners get fully engaged in developing interactions and it is possible

to notice the role of some discourse, pragmatic and interactional phenomena in facilitating and sustaining their conversation.

Following from the above discussion, it seems justifiable to argue that second language learners perform the above language functions (Bachman & Palmer, 1996) and convey and negotiate meanings through spoken discourse, relying on the complex interplay between discourse, pragmatic and interactional features. Therefore, in my view, "conversational competence" in a second language seems to comprise three major interrelated facets:

1. The management of discourse: mediating segments of discourse
2. The negotiation of illocutionary meaning: producing, manipulating and interpreting the pragmatic force of illocutionary acts
3. The deployment of conversational practices: showing an orientation to turn and sequence organisation and preference structure.

I will now attempt to justify this interplay by addressing each of the above facets and by detailing their key phenomena. Let us initially focus on the management of discourse. As previously mentioned in section 1.1, speakers express propositions through their utterances. Thus, more competent learners are able to produce and sustain longer stretches of discourse by means of cohesive devices, which signpost the direction of their utterances and facilitate interlocutors' understanding. They also rely on co-reference, substitution, ellipsis and lexical relationships in order to convey other cohesive relations within their own discourse. They are able to produce coherent discourse by adhering to thematic and information structure, which can be strengthened by such cohesive relations. They pay attention to listeners' response tokens, i.e. backchannels, and adjust their discourse accordingly. Furthermore, they are capable of employing discourse markers in order to launch and conclude topics, to signal relationships of sequence by indicating explicitly the order in which things occur or the organisation of the different segments, to monitor and manage the ongoing discourse by commenting explicitly on the process of talking itself , to indicate that they have not selected the most appropriate way of expressing things and that they are adding to or refining what they say, to signal that they are sensitive to listeners' needs and that they are monitoring the state of shared knowledge and to mark their stance or attitude towards the message.

Let us now address the negotiation of illocutionary meaning. As previously mentioned in section 1.2.1, speakers perform different acts through their utterances, in addition to expressing propositions. Thus, more competent learners are able to produce illocutionary acts according to the intended illocutionary force. In addition to pragmalinguistic knowledge (i.e. knowledge of linguistic forms to convey pragmatic meaning), they also rely on sociopragmatic knowledge to filter the appropriateness of specific functional exponents to communicative situations or contexts. They also activate both types of knowledge in order to convey the desired degree of power and relative distance. As listeners, they are able to interpret the illocutionary force of utterances, including the understanding of indirect speech acts, based on pragmalinguistic and sociopragmatic knowledge, illocutionary force indicating devices and contextual clues.

As for the manipulation of linguistic forms so as to convey subtle shades of illocutionary meaning (e.g. a suggestion versus a command), more competent learners are able to employ positive and negative politeness strategies in order to minimise the degree of imposition of face-threatening acts and to produce face-saving acts. They are capable of using pragmatic force modifiers in order to soften or strengthen the force of their messages and to make concepts fuzzier or less fuzzy. Strengthening modifiers are used for emphatic purposes, for showing involvement and as positive politeness strategies. Mitigating modifiers (i.e. hedges) are used to convey degrees of epistemic stance, to avoid direct responsibility for knowledge based on evidence, as negative politeness strategies and as vague language.

Furthermore, speakers engage in conversations in various situational settings. Second language learners may have to interact in English with their peers, with their teachers, with more proficient speakers, with examiners, with professionals, with people in the street, etc. The possible scenarios are countless. When interacting with people from the same nationality, learners are likely to transfer their pragmatic norms to the situation they are in. When interacting with native speakers of English or speakers of other languages, more competent learners are able to negotiate illocutionary meaning despite cultural differences.

Lastly, let us detail the third facet of conversational competence, namely the deployment of conversational practices. As previously mentioned in section 1.3, in face-to-face conversations, speakers establish and maintain social relationships and negotiate the exchange

of turns. More competent learners are usually able to carry out small talk as a means to initiate a conversation. When they initiate new turns, they usually display an orientation to transitional-relevant places. When they produce overlapping talk, this may occur as an effort to orient themselves to the rules of turn-taking. As coherence is achieved by contiguity, they are able to construct sequences of utterances by reference to the practices of adjacency pair organisation (including conversational routines), monitoring for action. They show an orientation to preference structure and usually deploy pre-sequences as a means to avoid dispreferred responses. They may include insert sequences as a means to remedy problems in hearing or understanding the preceding talk via repair. They may engage in post-expansion also for repair and for closing sequences. As listeners, they are able to employ backchannels in order to indicate that they are following what is being said and also to inform how the message is being received. Backchannels are used to encourage speakers to continue their ongoing turn and to signal convergence, engagement and information receipt. Speakers then engage in subsequent actions based on such feedback.

A cautionary note to add is that the above components are by no means mutually exclusive. There may be some overlap of specific phenomena within components due to their pragmatic functions. Discourse markers, for instance, are traditionally associated with discourse studies and then would normally be placed within the management of discourse. Nonetheless, they clearly play a role in the deployment of conversational practices. Thus, it is possible to state that more competent learners may use discourse markers to open, conclude or temporarily close a conversation and re-open previously closed or interrupted conversations when managing interactions. Likewise, backchannels are used by listeners to give feedback to speakers, and thus, are part of conversational practices. However, speakers also rely on such feedback to adjust their own discourse. Moreover, as expressing propositions is also as a sub-type of speech act, pragmatic force modifiers play a role in both discourse and illocutionary dimensions.

A second point to add is that learners can be more advanced in specific competences and less advanced in others. Some students who are keen on transmitting information during conversations tend to produce longer stretches of discourse. Other learners are good conversationalists since they are good listeners too, i.e., they express their interest and understanding or lack of understanding. In such circumstances, conversations are usually

symmetrical and we are able to perceive a sense of flow. Some students are particularly good at manipulating linguistic forms so as to convey different shades of illocutionary force while others tend to rely on more direct ones.

To conclude, in this section I have presented a revised view of conversational competence within a pragmatic domain, grounded on the notion of "meaning in interaction" (Thomas, 1995) and detailed the interplay between key discourse, pragmatic and interactional phenomena which comprise conversational competence in L2. Thus, I have proposed a revised model for conversational competence consisting of three interrelated facets:

1. The management of discourse includes a learner´s ability to:

- produce and sustain (longer) stretches of discourse, including the use of cohesive devices , co-reference, substitution, ellipsis and lexical relationships in order to convey cohesive relations;
- produce coherent discourse by adhering to thematic and information structure, which may be strengthened by cohesive relations;
- employ discourse markers to launch and conclude topics, to signal relationships of sequence, to monitor and manage the ongoing discourse, to indicate that they have not selected the most appropriate way of expressing things and that they are adding to or refining what they say, to signal that they are sensitive to listeners' needs and that they are monitoring the state of shared knowledge and to mark their stance or attitude towards the message.

2. The negotiation of illocutionary meaning includes a learner´s ability to:

- produce illocutionary acts according to the intended illocutionary force, relying on both pragmalinguistic and sociopragmatic knowledge and conveying the desired degree of power and relative distance;
- employ positive and negative politeness strategies in order to minimise the degree of imposition of face-threatening acts and to produce face-saving acts;
- use pragmatic force modifiers in order to soften or strengthen the force of their messages and to make concepts fuzzier or less fuzzy;

- interpret the illocutionary force of utterances, including the understanding of indirect speech acts, based on pragmalinguistic and sociopragmatic knowledge, illocutionary force indicating devices and contextual clues.

3. The deployment of conversational practices includes a learner´s ability to:

- carry out small talk as a means to initiate a conversation, including conversational routines;
- orient themselves to the rules of turn-taking;
- construct sequences of utterances by reference to the practices of adjacency pair organisation;
- show an orientation to preference structure;
- employ backchannels in order to indicate that they are following what is being said and also to inform how the message is being received.

By presenting the above conversational competence model and detailing its key discourse, pragmatic and interactional phenomena, it is possible to say that the first two objectives of this thesis have been accomplished. The third and final objective of this study is to investigate some of the key discourse, pragmatic and interactional phenomena comprising conversational competence in the oral production of twenty Brazilian learners of English at CEFR B1. Therefore, I have chosen to analyse the functions of pragmatic markers which lend themselves to corpus investigation and seem to be prototypical of the three proposed facets of conversational competence (despite some possible overlap), namely discourse marking adverbs, hedges and backchannels. Before concluding this chapter and moving on to unveil the methodology and the empirical data of this study, let us attempt to relate this revised view of conversational competence to the level descriptors of the "Common European Framework of Reference for Languages". While the Common European Framework of Reference for Languages is based on six broad reference levels (A1, A2, B1, B2, C1 and C2), the target level for my empirical investigations is B1, to be detailed and justified in the next sub-section.

2.4. THE COMMON EUROPEAN FRAMEWORK OF REFERENCE FOR LANGUAGES

The Common European Framework of Reference for Languages (CEFR so forth) is an extremely influential document since it provides a common basis for syllabus and

examination design, curriculum guidelines and textbooks across Europe and in different parts of the world. It details what abilities language learners have to learn and to develop in order to be able to use language for communicative purposes and to act effectively (Council of Europe, 2001). It defines levels of proficiency which enable learners' progress to be measured at each stage of the learning process and on a life-long basis. The uses of the CEFR include the planning of the language learning programme, the planning of language certification and the planning of self-directed learning. Learning programmes and certification can be global, i.e. bringing learners forward in all dimensions of language proficiency and communicative competence; modular, i.e. improving learners' proficiency in a restricted area for particular purposes; weighted, i.e. emphasising learning in certain directions and producing a profile in which a higher level is attained in some areas of knowledge and skill than others; and partial, i.e. taking responsibility only for certain activities and skills and leaving others aside. Thus, "in order to fulfil its functions, such a Common European Framework must be comprehensive, transparent and coherent" (Council of Europe, 2001, p. 7)

The CEFR is an action-oriented approach to language teaching and learning. Such an approach views language users and learners as social agents who perform tasks in specific circumstances, environments and within particular fields of actions. It also takes into account "the cognitive, emotional, volitional resources and the full range of abilities specific to and applied by the individual as a social agent" (Council of Europe, 2001, p. 9). Language use and learning are described as follows:

> Language use, embracing language learning, comprises the actions performed by persons who as individuals and as social agents develop a range of competences, both general and in particular communicative language competences. They draw on the competences at their disposal in various contexts under various conditions and under various constraints to engage in language activities involving language processes to produce and/ or receive texts in relation to themes in specific domains, activating those strategies which seem most appropriate for carrying out the tasks to be accomplished. The monitoring of these actions by the participants leads to the reinforcement or modification of their competences. (Council of Europe, 2001, p. 9)

Milanovic and Weir (2010) review the impact of the CEFR on education in Europe. As for advantages, Milanovic and Weir (2010) state that the framework has detailed learning objectives, raised awareness of language issues and also informed researchers, policy makers, assessment providers and teachers. Nonetheless, Milanovic and Weir (2010) stress that there

is some concern that the CEFR has been adopted or interpreted as a fixed standard. Over time, test providers and examination boards have been highly recommended to link their examinations to the CEFR levels by government agencies. Initially, the Council of Europe provided a "toolkit" of resources to inform and facilitate the process of aligning tests to the framework. This toolkit included a draft pilot manual for relating language examinations to the CEFR, a technical reference supplement and a revised version of the Manual. The Council also set up forums where practitioners were invited to reflect upon and share their experiences.

Milanovic and Weir (2010) explain that "The Association of Language Testers" (ALTE) and Cambridge English Language Assessment (a founder member of ALTE) have supported the development of the toolkit, which is intended to equip users to make good use of the framework vis-à-vis their own objectives. Both organisations have been directly involved in the following projects: co-ordinating the development of a User's Guide for Examiners (1996), developing the EAQUALS/ALTE European Language Portfolio (ELP), both in hard copy and electronic forms (from 2000), providing support for the authoring and piloting of the draft Manual for relating examinations (since 2002/3), contributing to benchmarking materials and examples of test items and tasks to accompany the CEFR (as from 2004) and developing content analysis grids for speaking and writing materials. In addition, Cambridge English Language Assessment has been coordinating the development of Reference Levels Descriptions for English – "The English Profile Programme" since 2005. Milanovic and Weir (2010, p. ix) highlight that "the Profile in English will contribute significantly to the usefulness of the CEFR as a practical tool".

Furthermore, from the early 1990's, Cambridge English Language Assessment together with ALTE worked to develop an empirically derived common scale level as well as the comparison of examinations across languages on the ALTE 5-level system. "The Common Scale ... relates different testing events within a single frame of reference, greatly facilitating the development and consistent application of standards" (Milanovic and Weir, 2010, p. x). Since the introduction of the Common Scale, a large number of candidates (in millions) have taken the full range of Cambridge English examinations and their responses have provided data for further refinement of the scales.

Milanovic and Weir (2010, p. x) clarify that "the CEFR itself is deliberately underspecified and incomplete", which makes it an appropriate tool for comparisons of practices across different contexts in Europe and beyond. While it is useful as a common framework with six broad reference levels, it is not applicable to all contexts, without user intervention so as to customise its suitability. North, Martyniuk and Panthier (2010) point out that the Committee of Ministers representing the 47 members states of the Council of Europe adopted the "Recommendation CM/Rec (2008) 7" on the use of the CEFR in July 2008. This recommendation included the careful consideration of very important principles underlying the coherent, realistic and responsible use of the CEFR (North et al., 2010, p. 13):

a) The CEFR is purely descriptive- not prescriptive or normative;

b) The CEFR is language neutral - it needs to be applied and interpreted appropriately with regard to each specific language;

c) The CEFR is context neutral - it needs to be applied and interpreted appropriately with regard to specific educational context in accordance with the needs and priorities specific to that context;

d) The CEFR attempts to be comprehensive, in that no aspects of language knowledge, skills and use are deliberately left out of consideration. It cannot, of course, claim to be exhaustive. Further elaboration and developments are welcome.

e) The CEFR offers a common language point of reference as a basis for stakeholders to reflect upon and critically analyse their existing practice and to allow them to better "situate their efforts" in relation to one another.

f) The use of the CEFR should contribute to increased transparency of processes and procedures, improved quality of provision and comparability of outcomes.

g) The use of the CEFR should contribute to the promotion of the basic educational values for which the Council of Europe stands, such as social inclusion, intercultural dialogue, active democratic citizenship, language diversity, plurilingualism, learner autonomy and lifelong learning.

With regards to assessment, Weir (2005) clarifies that the CEFR is not sufficiently comprehensive, coherent or transparent for uncritical use in language testing for two reasons. Firstly, the descriptor scales do not suffice to account for how variation in terms of contextual parameters may affect performances by raising or lowering the actual level of difficulty related to the "Can Do" statement target. Secondly, the cognitive validity of a test must be

addressed by any specification on which a test is based. Weir indicates that the CEFR should not be seen as a prescriptive device but as a heuristic one, which can be refined and developed so as to meet specific aims.

In this section, we have discussed the general features of the Common European Framework of Reference for Languages and its impact on education. The CEFR provides a common basis for syllabus and examinations design, curriculum guidelines and textbooks since it details what abilities language learners have to learn and to develop in order to be able to use language for communicative purposes and to act effectively (Council of Europe, 2001). It is an action-oriented approach to language teaching and learning. While it is useful as a common framework with six broad reference levels, it is not applicable to all contexts, without user intervention so as to customise its suitability (Milanovic & Weir, 2010).

Let us now address the level descriptors of the Common European Framework of Reference for Languages.

2.4.1. The CEFR levels

One of the objectives of the CEFR is to describe the levels of proficiency required by existing standards, tests and examinations so as to make comparisons between different systems of qualifications possible. As a result, a descriptive scheme and the CEFR levels were developed, guided by four criteria. Firstly, "a common framework scale should be context-free in order to accommodate generalisable results from different specific contexts" (Council of Europe, 2001, p. 21). Secondly, such a description needs to be based on theories of language competence. However, "the available theory and research is inadequate to provide a basis for such a description" (Council of Europe, 2001, p. 21). Having said that, the categorisation and description needs to be theoretically grounded and at the same time remain user-friendly. Thirdly, the points on the scale at which specific activities and competences are situated in a common framework should be objectively determined since they are based on a theory of measurement (Council of Europe, 2001, p. 21). Lastly, the number of levels adopted should be adequate to show a progression in different sectors, but, in any particular context, should not exceed the number of levels between which consistent distinctions can be made (Council of Europe, 2001, p. 21).

Let us initially examine how communicative competence is viewed in the CEFR. In addition to general capacities, users/ learners rely on a more specifically language-related communicative competence[24] consisting of three major components: linguistic, sociolinguistic and pragmatic competences (Council of Europe, 2001, p. 108). Linguistic competences include lexical, phonological, syntactical knowledge and skills, independently of the sociolinguistic value of its variations and of the pragmatic functions of its realisations. This component also relates to cognitive organisation, to the way this knowledge is stored and to its accessibility. Sociolinguistic competences relate to the sociocultural conditions of language use such as rules of politeness, norms governing relations between generations, genders, classes and social groups and linguistic codification of certain fundamental rituals in the functioning of a community. Pragmatic competences refer to the functional use of linguistic resources (production of language functions and speech acts), drawing on scenarios or scripts of interactional exchanges. This component also includes the mastery of discourse, cohesion and coherence, the identification of text types and forms, irony and parody.

Within pragmatic competences, the notion of "conversational competence" is briefly mentioned under functional competence, which refers to the use of both spoken and written texts in communication for particular functional purposes, including microfunctions, macrofunctions and interaction schemata (Council of Europe, 2001, p. 125). Microfunctions relate to categories for the functional use of single (usually short) utterances, usually as turns in an interaction (e.g. identifying and suggesting) while macrofunctions refer to categories for the functional use of spoken discourse or written text consisting of a (sometimes extended) sequence of sentences (e.g. description and narration). Interaction schemata encompass knowledge of and ability to use patterns of social interaction which underlie communication, such as verbal exchange patterns (e.g. question/answer and statement/agreement-disagreement). Discourse competence is described as a pragmatic sub-competence which refers to the ability of a user/learner to arrange sentences in sequence so as to produce coherent stretches of language (Council of Europe, 2001, p. 123). It includes knowledge of and ability to control the ordering of sentences in terms of thematic development, coherence and cohesion and turn-taking, among others elements.

[24] For a full description of communicative language competence in the CEFR, please see Council of Europe (2001, pp. 108-130).

For the purposes of this thesis, I decided not to include the above view of communicative competence in my previous critical review (section 2.3) for two main reasons. Firstly, the framework is "deliberately underspecified and incomplete" (Milanovic and Weir, 2010, p. x). As a result, its common reference points and the wording of descriptors are likely to develop over time as experience of member states and of institutions with related expertise is incorporated in the description (Council of Europe, 2001, pp. 23-24). Secondly, I chose to look at the above description of communicative competence only after I had refined my own construct of conversational competence. I have been a speaking examiner for different CEFR levels for 20 years and I am thoroughly familiar with assessment scales. Thus, I wanted to propose a fresh and unbiased view of conversational competence within a purely pragmatic domain rather than as part of a broader communicative competence framework within assessment realms. Nonetheless, I was particularly pleased to see that in the above construct, both discourse competence and functional competence are regarded as pragmatic competences, corroborating my assumption that a conversational competence construct should be placed within a pragmatic domain.

Let us now focus on the CEFR levels themselves. The proposed scheme of levels adopts a hypertext branching principle, starting from an initial division into three broad levels - A, B and C (Council of Europe, 2001, p. 23):

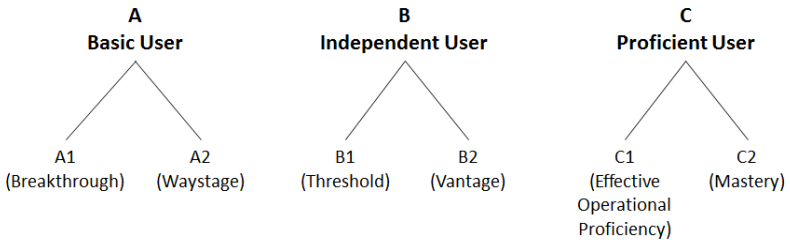


Figure 3: CEFR levels

Within the EFL tradition, category A comprises elementary (A1) and pre-intermediate (A2) levels, category B includes intermediate (B1) and upper-intermediate (B2) levels and category C encompasses advanced (C1) and proficiency (C2) levels. North, Martyniuk and Panthier (2010, pp. 4-5) highlight the most salient features of the CEFR levels:

a) At level A1, learners are able to:

- interact in a simple way, ask and answer simple questions about themselves, where they live, people they know, and things they have; initiate and respond to simple statements in areas of immediate need or in very familiar topics, rather than relying purely on a rehearsed repertoire of (tourist) phrases.

b) Level A2 reflects the Waystage[25] specification with:

- the majority of descriptors stating social functions: greet people, ask how they are and react to news; handle very short social exchanges; ask and answer questions about what they do at work and in free time; make and respond to invitations; discuss what to do, where to go and make arrangements to meet; make and accept offers;
- and descriptors on getting out and about: make simple transactions in shops, post offices or banks; get simple information about travel; ask for and provide everyday goods and services.

c) Level B1 reflects the Threshold Level with two particular features:

- maintaining interaction and getting across what you want to communicate: give and seek personal views and opinions in an informal discussion with friends; express the main point he/she wants to make comprehensibly; keep going comprehensibly, even though pausing for grammatical and lexical planning and repair are very evident, especially in longer stretches of free production;
- coping flexibly with problems in everyday life: deal with most situations likely to arise when making travel arrangements through an agent or when actually travelling; enter unprepared into conversations on familiar topics; make a complaint.

d) Level B2 reflects three new emphases:

- effective argument: account for and sustain opinions in discussion by providing relevant explanations, arguments and comments; explain a viewpoint on a topical issue giving the advantages and disadvantages of various options;
- holding your own in social discourse: interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without imposing strain on either party; adjust to the changes of direction, style and emphasis normally found in conversation;

[25] Waystage and Threshold refer to previous Council of Europe levels presented at the 1977 Ludwigshafen symposion (North et al., 2010, p. 4).

- displaying a new degree of language awareness: correct mistakes if they have led to misunderstandings; make a note of "favourite mistakes" and consciously monitor speech for them.

e) Level C1 is characterised by access to a broad range of language that results in fluent, spontaneous communication:

- expressing oneself fluently and spontaneously, almost effortlessly; having a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions; there is little obvious searching for expressions or avoidance strategies – only a conceptually difficult subject can hinder a natural smooth flow of language;
- producing clear, smoothly flowing, well-structured speech, showing controlled use of organisational patterns, connectors and cohesive devices.

f) Level C2 represents the degree of precision and naturalness typical of highly successful learners:

- conveying finer shades of meaning precisely by using, with reasonable accuracy, a wide range of modification devices;
- having a good command of idiomatic expressions and colloquialisms with awareness of connotative level of meaning.

The description of the salient features of the CEFR levels enables us to relate the different facets of the conversational competence model proposed in the previous sub-section as from level A1 already. While the ability to "ask and answer simple questions" seems to relate to the management of discourse since the focus is on the transmission of information, the ability to "initiate and respond to simple statements" already points to conversational practices, when one considers the practices of pair-adjacency and backchanelling. At level A2, the deployment of conversational practices is also very salient as learners are likely to use conversational routines for performing social functions like greetings and to show an orientation to preference structure when making and accepting offers. The negotiation of illocutionary meaning is also present at A2, since learners are expected to perform functions for transactional purposes such as transactions in shops, post offices or banks, etc.

At level B1, one can more clearly notice the interplay between discourse, pragmatic and interactional phenomena. In order to "maintain interaction", learners may initially express their views producing extended discourse, relying on features such as cohesive and reference devices and discourse markers. At the same time, they are also likely to engage in interactional practices like turn-taking routines and backchannels so as to keep the interaction going. "Getting across what you want to communicate" is a really broad pragmatic function as it can be applied to the three facets. The notion of "getting across" is closely related to the negotiation of illocutionary meaning, since learners are likely to rely on pragmalinguistic and sociopragmatic knowledge and contextual clues in order to produce and interpret speech acts. More competent learners may also employ politeness strategies and hedges to manipulate the force of illocutionary acts. As listeners, more competent learners are likely to use backchannels to indicate how the message is being received while as speakers, they may adjust their discourse based on such feedback, and thus, manage both interactional practices and discourse respectively. "Entering unprepared into conversations on familiar topics" draws on the three facets of conversational competence being a superordinate function.

At level B2, the abilities "accounting for and sustaining opinions" and "expressing a viewpoint on a topical issue" initially require the activation of discourse features so that learners are capable of producing coherent and extended discourse. Learners are also likely to rely on discourse markers so as to mark their stance to the message. In addition, they may use pragmatic force modifiers for showing involvement, for conveying degrees of epistemic stance and for avoiding direct responsibility for knowledge based on evidence, and thus, manipulating illocutionary meaning. "Holding your own image in social discourse through interaction and adjustments and self-repair" seems to be closely related to the deployment of conversational practices since learners are expected to show an orientation to turn-taking rules and engage in repair sequences when performing such a broad function. Nonetheless, more competent learners are also likely to manipulate illocutionary meaning when employing positive and negative politeness strategies so as to produce face-saving acts and to minimise the degree of imposition of face-threatening acts.

At level C1, "producing clear, smoothly flowing, well-structured speech, showing controlled use of organisational patterns, connectors and cohesive devices" strengthens the role of discourse features in the management of discourse. "Expressing oneself fluently and

spontaneously, almost effortlessly" and "having a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions" points to the mastery of code components such as grammatical and lexical knowledge. At level C2, "conveying finer shades of meaning precisely by using, with reasonable accuracy, a wide range of modification devices" requires the ability to manipulate illocutionary meaning through the use of pragmatic force modifiers and politeness strategies.

In order to pursue the objectives of my research questions, I chose to focus my empirical investigation on level B1 for three main reasons. Firstly, CEFR B1 is also referred to as the intermediate level or the intermediate plateau within EFL contexts. Learners who have reached this level do not usually sense progress as markedly as before. In addition, moving beyond this plateau to the upper-intermediate and advanced levels is sometimes frustrating and, on some occasions, unattainable. Secondly, as illustrated in the previous discussion, at B1 learners are likely to rely on the interplay of some of the key discourse, pragmatic and interactional phenomena highlighted in the proposed model of conversational competence. Thirdly, the Brazilian programme "Science without Borders" requires Brazilian undergraduate students to sit formal language examinations. In order to be elegible to study at an overseas university for a year, applicants need to achieve a successful performance at CEFR B2 and above, depending on the demands of the target country institution and on the examination taken per se. Therefore, I hope that the findings of this study may shed some light on the intermediate plauteau phenomenon, providing teachers with authentic samples of Brazilian learners' strengths and weaknesses in terms of spoken production, which may be considered when preparing other students for the speaking component of international examinations. The methodology and instruments employed in this empirical investigation will be detailed in the next chapter.

To sum up, in this section I have presented the CEFR view on communicative language ability and described the CEFR levels. I have also related the most salient features of the CEFR levels to the different facets of the proposed model for conversational competence, highlighting the interplay of discourse, pragmatic and interactional phenomena within specific functions. Lastly, I have justified theoretical and methodological choices regarding the proposed model for conversational competence and the target level for empirical investigation, namely CEFR B1. Let us finalise this chapter by briefly describing the English

Profile Programme, whose purpose is to locate the acquisition of specific linguistic features of the English language by second language learners at specific CEFR levels.

2.4.2. The English Profile

According to Milanovic and Weir (2010), ALTE's testing systems have developed alongside the CEFR over the last two decades. Many of these developments can provide rich data and analysis to refine the CEFR as they relate to a number of European languages. The English Profile Programme (EPP) is intended to transpose the CEFR into English so that it immediately informs curriculum designers, teachers and assessment organisations, among others. Similar projects are being developed for other languages, in countries where ALTE has working partners. One of the main aims of the EP project is to analyse learner language so as to detail what learners of English can or cannot do at different levels and to address how well learners perform using linguistic exponents of the language at their disposal. These analyses are based on authentic learner input available in the Cambridge Learner Corpus, which contains both written and spoken data ranging from levels A1 to C2 of the CEFR. Milanovic and Weir (2010, p. xvi) conclude that "Cambridge ESOL is now in a position to begin a systematic and empirically based approach to specifying more precisely how the CEFR can be operationalised in English, and this in turn will lead to better and more comprehensive illustrative descriptors".

Hawkins and Filipovic (2012, pp. 112-133) report on the empirical findings of the English Profile Programme so far. They present a summary of a representative set of positive linguistic features, i.e. "correct linguistic properties that have been acquired as transitional features at different CEFR levels and that generally persist at higher levels" (Hawkins & Filipovic, 2012, p. 112). Findings[26] indicate that learners at A2 are able to produce intransitive and transitive clauses, verbs with a finite complement clause, verbs with an infinitival complement, direct WH-questions, pronouns plus infinitive and ditransitive clauses (e.g. *"I can give you my guitar*"). At B1, learners can produce postnominal modification with *–ed* and *– ing*, verbs with infinitival complements, verbs with *–ing* complements, *It* Extraposition with finite clauses (e.g. "*It's true that I don't need a ring to make me remember you*"), genitive relatives (e.g. "*I met a very nice boy whose name's John*"), indirect WH-questions in finite

[26] For a full account of linguistic features at CEFR levels, please see Hawkins and Filipovic (2012, pp. 103-165).

clauses, indirect WH-questions in infinitival phrases and complex auxiliaries (e.g. "*would rather*", "*had better*"). Lastly, at B2 learners are capable of producing adverbial subordinate clauses with –ing (e.g. "*Talking about spare time, I think we could go to the Art Museum*"), pseudocleft (e.g. "*What interests me is*…") and verbs with an NP plus finite complement clause (e.g. "*I told him I loved his songs*").

As for modality[27], Hawkins and Filipovic (2012: 131-133) detail criterial features involving the modal auxiliaries *may*, *might*, *can*, *must* and *should* in the Cambridge Learner Corpus with their respective meanings:

1. **May**

   A2 Possibility (Epistemic): "*Then we may go sightseeing*".

   B1 Permission (Deontic): "*May I suggest that you book me in for the new accommodation*?"

2. **Might**

   A2 Possibility (Epistemic): "…*the paint might make our T-shirts dirty*".

   C1 Permission (Deontic): "*Might I tell you what we discuss?*"

3. **Can**

   A1 Ability implicating Permission (Deontic), see Searle (1975): "*So please, can you make me a big salad?"*

   A2 Possibility (Epistemic): "*We can meet at our school*".

4. **Must**

   A2 Obligation (Deontic): "*We must be there at 7 o'clock in the morning*".

   B1 Necessity (Epistemic): "*She must be feeling so happy!"*

5. **Should**

   A2 Advice (Deontic): "*You should wear old clothes because we will get dirty*".

   B1 Probability (Epistemic): "*I have invited all his friends, so we should be 28 people*".

[27] Epistemic modality refers to speakers' expression of judgements about the factual status of propositions while deontic modality is traditionally associated with the notions of permission and obligation.

In addition to the above lists of linguistic features, the English Profile Programme has a dedicated website (http://www.englishprofile.org/) which invites researchers to contribute to the project and includes a free subscription to a sub-division of the project named "the English Profile Vocabulary" (http://www.englishprofile.org/index.php/wordlists/free-subscription). Once registered on the website, it is possible to type in specific lexical items and locate them at specific CEFR levels. For instance, the adverb "*actually*" first appears in the Cambridge Learner Corpus at A2, when used for emphasising what is really true (e.g. "*I don't actually like seafood*"). It also appears at B1 when used for emphasising something surprising (e.g. "*A friend of mine from university was actually at the same concert*") and at B2 when used for correcting someone (e.g "*Actually, Gavin, it was Tuesday of last week, not Wednesday*").

Concluding, in this sub-section I have described the English Profile Programme and listed a representative set of positive linguistic features (Hawkins & Filipovic, 2012). The EPP analyses learner language so as to detail what learners of English can or cannot do at different levels, based on authentic learner input available in the Cambridge Learner Corpus. It contains both written and spoken data ranging from levels A1 to C2 and accounts for linguistic features like grammatical items, syntantic structures, modal auxiliary verbs (including specific uses) and lexical items (including specific meanings). Modality data will be further referred to in chapter 5.

SUMMARY

In this chapter, I have addressed Interlanguage Pragmatics, critically revisited competence frameworks and proposed a conversational competence model in L2 within a pragmatic domain. Interlanguage Pragmatics is a second-generation hybrid belonging to the interdisciplinary fields "Second Language Acquisition" and "Pragmatics" (Kasper & Blum-Kulka, 1993). ILP within the SLA domain examines how L2 learners develop the ability to understand and perform action in a target language whereas, as a sub-set of Pragmatics, it investigates how non-native speakers comprehend and produce action (Kasper & Rose, 2002). ILP investigation focuses on non-native pragmatic comprehension, the production of linguistic action, pragmatic development, pragmatic transfer and pragmatic failure (Kasper & Blum-Kulka, 1993).

Schmidt (1993) and Bialystok (1993) are highly influential cognitive theoreticians who account for pragmatic learning. Corsetti's results (2009) are aligned with Schmidt's (1993) claim that close attention to the relevant features of input facilitates pragmatic learning. Negative transfer is the influence of L1 pragmatic competence on interlanguage pragmatic knowledge that differs from the L2 target (Kasper & Blum-Kulka, 1993). Positive transfer refers to pragmatic behaviours or other knowledge displays that are consistent across L1, interlanguage and L2 and that usually result in communicative success (Kasper & Blum-Kulka, 1993).

Dell Hymes's (1972) concept of communicative competence and the subsequent frameworks by Canale and Swain (1980), Canale (1983) and Bachman (1990) contributed to a shift in L2 pedagogy towards communicative language learning and teaching, which in turn, resulted in an increased interest in pragmatic studies focusing on language learners in the mid 1980s (Schauer, 2009). In Canale and Swain (1980) and Canale's (1983) view, comunicative competence consists of four sub-competences: linguistic or grammatical competence, sociolinguistic competence, strategic competence and discourse competence. It was argued that their model fails to indicate how the different components interact with one another (Csépes, 2009).

In Bachman's (1990) interactive framework, communicative language ability includes three components: language competence, strategic competence and psychophysiological mechanisms. Language competence comprises two major competences consisting of several categories that interact with one another and also with features of the language use situation: organisational competence (grammatical and textual competences) and pragmatic competence (illocutionary and sociolinguistic competences). While Bachman's original model (1990) represents a major reconceptualisation since it locates competence within a wider performance framework (Csépes, 2009), it fails to account for the co-constructed nature of communicative performance (McNamara, 1997).

Both views on communicative competence place the notion of conversational competence within a broader framework, which accounts for both oral and written discourse, namely discourse competence (Canale, 1983) and textual competence (Bachman, 1990). Nonetheless, the ability to produce spoken discourse and interact with other speakers seems complex

enough to justify a separate characterisation from the ability to produce written discourse, especially when one considers the negotiation of meaning which takes place in real time. Thus, different characterisations of interactional and conversational competence were presented.

Interactional competence was described in terms of extended routines, the coherence and the use of speech markers and the use of rhetorical features while conversational competence was detailed in terms of sequence organisation, conversational style and turn-taking routines (Schmidt, 1983). Other views on conversational competence include speakers' knowledge of how speech acts are used in social situations (Richards & Sukwiwat, 1985) and the subset of linguistic and strategic competencies that are implicated in conversation and that draw on both textual and pragmatic knowledge (Thornbury & Slade, 2006). Conversational-interactional competence was defined as speakers´ ability to produce illocutionary acts conveying the intended illocutionary force as well as the ability to manage dialogic and monologic discourse (Corsetti, 2009). While the above characterisations of conversational competence include elements derived from Discourse, Pragmatics and Conversation Analysis, they do not specify the interplay between discourse, pragmatic and interactional phenomena or account for both comprehension and production dimensions.

Thus, I argued that L2 learners perform language functions (Bachman & Palmer, 1996) and convey and negotiate meanings through spoken discourse, relying on the complex interplay between discourse, pragmatic and interactional features. I proposed a three-part model of conversational competence within a pragmatic domain, grounded on two assumptions: conversation is prototypical of language usage (Levinson, 1983) and language in use involves the negotiation of meaning (Thomas, 1985). The management of discourse includes a learner's ability to: produce and sustain (longer) stretches of discourse, including the use of cohesive devices , co-reference, substitution, ellipsis and lexical relationships; produce coherent discourse by adhering to thematic and information structure; employ discourse markers to perform a number of organisational and pragmatic functions (as detailed on page 105). The negotiation of illocutionary meaning includes a learner's ability to: produce illocutionary acts according to the intended illocutionary force; employ positive and negative politeness strategies in order to minimise the degree of imposition of face-threatening acts and to produce face-saving acts; use pragmatic force modifiers in order to soften or strengthen the

force of their messages; interpret the illocutionary force of utterances, including the understanding of indirect speech acts. The deployment of conversational practices includes a learner's ability to: carry out small talk as a means to initiate a conversation, including conversational routines; orient themselves to the rules of turn-taking; construct sequences of utterances by reference to the practices of adjacency pair organisation; show an orientation to preference structure; employ backchannels in order to indicate that they are following what is being said and also to inform how the message is being received.

In the final part of chapter 2, I related the different facets of the proposed model for conversational competence to the CEFR level descriptors: A1 (elementary), A2 (pre-intermediate), B1 (intermediate), B2 (upper-intermediate), C1 (advanced) and C2 (proficiency). While individual components are activated as from level A1, the interplay between discourse, pragmatic and interactional phenomena is more noticeable as from level B1. I chose CEFR level B1 for empirical investigation, based on the notion of the intermediate plateau and on the requirements of Brazilian programme "Science without Borders". Furthermore, the project "English Profile" locates the acquisition of specific linguistic features of the English language by L2 learners at specific CEFR levels. A summary of a representative set of positive linguistic features was presented, detailing correct linguistic properties which have been acquired as transitional features at different CEFR levels and which generally persist at higher levels (Hawkins & Filipovic, 2012). The modal auxiliaries "*may"*, "*might"*, "*can"*, "*must"* and "*should"* with their epistemic and deontic meanings were also matched against the CEFR levels, and will be further referred to in chapter 5.

All things considered, it is possible to say that the first two objectives of this thesis have been accomplished. Let us now move on to address its final aim. In the next chapter, I initially detail the foundations of Corpus Linguistics and then describe the methodology employed in my empirical investigations. In order to illustrate the different facets of the proposed model for conversational competence and pursue my research questions, I focus on discourse marking adverbs as a means to mediate segments of discourse in chapter 4, I analyse the role of hedges in the negotiation of illocutionary meaning, namely the mitigation of representative speech acts, in chapter 5 and I address the pragmatic functions of backchannels as a means to express good listenership in chapter 6.

3. METHODOLOGY

In the previous chapter, I endeavoured to accomplish the theoretical aims of this thesis, namely revisiting the construct "conversational competence" in L2 within a pragmatic domain and detailing its key discourse, pragmatic and interactional phenomena. The final objective of this study has an empirical nature since it aims at investigating some of these key phenomena in the oral production of twenty Brazilian learners of English at CEFR B1. Thus, based on empirical data, I address four research questions which are pursued through more specific sub-questions in subsequent chapters:

RQ1: *How do Brazilian learners at CEFR B1 mediate segments of discourse in conversations?*
RSQ1: *What are the most common discourse marking adverbs used to mediate segments of discourse in conversations by Brazilian learners at CEFR B1? How do their frequency of use and pragmatic functions compare to those of native speakers?*

RQ2: *How do Brazilian learners at CEFR B1 mitigate representative speech acts?*
RSQ2: *What are the most common explicit and implicit adverbial hedges used to mitigate representative speech acts by Brazilian learners at CEFR B1? How do their frequency of use and pragmatic functions compare to those of native speakers?*

RQ3: *How do Brazilian learners at CEFR B1 express good listenership?*
RSQ3: *What are the most common minimal response tokens used to express good listenership by Brazilian learners at CEFR B1? How do their frequency of use and pragmatic functions compare to those of native speakers?*

RQ4: *What are the pragmatic markers which seem to pose a challenge to Brazilian learners of English at CEFR B1?*
RSQ4: *Based on the findings of research sub-questions 1 to 3, what are the pragmatic markers which seem to pose a challenge to Brazilian learners of English at CEFR B1?*

Chapter 3 unveils the methodology employed in the above empirical investigations. I initially present different views (Granger, 2002, Biber, Conrad and Reppen, 1998, McEnery & Hardie,

2012) on the status of Corpus Linguistics and address learner corpora, corpus-based investigations and techniques. Next, I discuss combined approaches to the investigation of spoken discourse, namely Corpus Pragmatics (Jucker, 2013) and CLCA (O'Keeffe and Walsh, 2012), which include corpus techniques. In the final part of this chapter, I describe the empirical study of this thesis, its subjects, instruments and corpora and the methodological choices and procedures employed for my empirical investigations.

3.1. CORPUS LINGUISTICS

There has been some pertinent debate about the status of Corpus Linguistics, whether it is best defined as a linguistic sub-field, as a methodology or neither. For Granger (2002), "Corpus Linguistics" is a linguistic methodology grounded on the use of electronic collections of naturally occurring texts (both oral and written), i.e. corpora. As a methodology, its main strength lies in the potential to change theories of language as it enables researchers to form, challenge and abandon hypotheses by investigating real and authentic samples of language in use. The combination of computer software tools with an impressive amount of language data has revealed previously unsuspected linguistic phenomena (Stubbs, 1996). Taking into consideration this potential to change theories of language, Corpus Linguistics may be neither a linguistic sub-field nor a methodology. It may be best viewed as a perspective since it is both an instrument and an approach to linguistics. A number of influential corpus linguists such as Douglas Biber favour the term "corpus-based approach".

For Biber, Conrad and Reppen (1998), the corpus-based approach emphasises language in use by investigating how users exploit the resources of their language, in naturally occurring texts. Therefore, corpus-based analyses comprise four key features: they are empirical (analysing patterns of use in natural contexts); they utilise a corpus as the basis for analysis; they make extensive use of computers for analysis, both automatic and interactive features; they depend on both quantitative and qualitative analytical techniques. Biber, Conrad and Reppen (1998) add that corpus-based analysis should be regarded as a complementary approach to more traditional ones such as lexicography and sociolinguistics, among many others. Research questions for corpus-based studies usually derive from other kinds of investigations, for instance, they can originate from prior structural studies; from a hypothesis or theoretical framework or from intuition and anecdotal evidence. Furthermore, detailed corpus-based

analyses can focus on specific linguistic features in particular texts, complementing the findings from analyses of larger corpora.

Following this, corpus-based investigations rely on a combination of quantitative and qualitative methods. In qualitative research, no attempt is made to assign frequencies to the linguistic features that are identified in the data (McEnery & Wilson, 1996, p. 62). The main purposes of qualitative research are to use data only as a basis for identifying and describing aspects of usage in the language and to provide real-life examples of particular phenomena. In quantitative research, on the other hand, features are classified, counted and more complex statistical models are built in an attempt to explain observed phenomena (McEnery & Wilson, 1996, p. 62). McEnery and Wilson (1996) acknowledge that a stage of qualitative research is usually a precursor for quantitative analysis as categories for classification must be first identified before linguistic phenomena can be classified and counted. However, the researchers suggest that it is more useful to consider them as two different but not necessarily incompatible perspectives on corpus data.

For McEnery and Wilson (1996, p. 62), qualitative forms of analysis offer a rich and detailed perspective on the data since both rare and more frequent phenomena can be accounted for, enabling fine distinctions to be drawn within specific varieties. The main disadvantage of qualitative analysis lies in the fact that such findings may not be extended to wider populations with the same degree of certainty with which quantitative analysis can. While a corpus may be statistically representative, it may not be possible to test specific findings of the research so as to discover whether they are statistically significant[28] or more likely to be due to chance. Quantitative analysis, on the other hand, allows its findings to be generalised to a larger population and it makes the direct comparison between different corpora possible. Nonetheless, the data derived from quantitative analysis is necessarily less rich than that obtained from qualitative analysis (McEnery & Wilson, 1996, p. 63).

Now that we have addressed the main differences between quantitive and qualitative methods, let us present a third view on the status of Corpus Linguistics. For McEnery and Hardie

[28] There are different significance tests available to corpus linguists such as the chi-squared test, the [Student's] t-test and Wilcoxon's rank test (McEnery & Wilson, 1996, p. 70). The chi-squared test, for instance, compares the difference between the actual frequencies which have been observed (the observed frequencies) with those one would expect if no factor other than chance had been operating to affect the frequencies (the expected frequencies).

(2012), corpus linguistics is an area that focuses upon a set of procedures and methods for studying language. Although some of its procedures are still in development, some of them are well-established and central to the approach such as concordancing. In their view, corpus-linguistics has the potential to reorient one's approach to the study of language, to refine and redefine theories of language and to enable the use of theories of language which were previously difficult to be explored without corpus-linguist methods. McEnery and Hardie (2012) add that whilst it is possible to make some generalisations that characterise much of what is understood as corpus linguistics, it is important to highlight that corpus linguistics is a heterogeneous field. There are differences within corpus linguistics that distinguish and sub-categorise varying approaches to the use of corpus data.

Tognini-Bonelli (2001) draws a distinction between two types of approaches to corpus investigations. In her view, a corpus-based approach refers to a methodology that uses the corpus so as to detail, test, validate or exemplify theories or descriptions that were formulated before large corpora become available to inform language study. Thus, the commitment to the data as a whole is not ultimately very strict or systematic. In a corpus-driven approach, on the other hand, linguists commit to the integrity of the data as a whole and provide comprehensive descriptions based on corpus-evidence. The corpus is regarded as more than a repository of examples to back pre-existing theories. In other words, "the theoretical statements are fully consistent with, and reflect directly, the evidence provided by the corpus" (Tognini-Bonelli, 2001, p. 84). McEnery and Hardie (2012) explain that "corpus-based research" is grounded on the definition of corpus linguistics as a method. Conversely, "corpus-driven linguistics" rejects this characterisation and claims that corpus investigations should generate hypotheses about language. McEnery and Hardie (2012, p. 6) reject the dichotomy corpus-based/corpus-driven as, in their view, a corpus does not have a theoretical status and thus, all corpus linguistics can justly be described as corpus-based.

Now that we have discussed the different approaches to Corpus Linguistics and detailed quantatitive and qualitative methods of investigation, let us focus on the second language learner more specifically by addressing learner corpora.

3.1.1. Learner corpora

According to Granger (2002), corpus-based research over the last thirty years has provided better descriptions of different registers (informal conversation, formal speech, journalese, academic writing, sports reporting, etc.) and dialects of native English (British English versus American English, male versus female language, etc.). However, it was not until the late 1980s and the early 1990 that corpora of non-native varieties, i.e. learner corpora, started to be collected. The investigation of learner corpora can inform both second language acquisition (SLA) research and foreign language teaching research. The number of SLA studies based on learner corpora has been traditionally small due to the fact that, until quite recently, data collection and analysis would be an incredibly time-consuming and highly demanding task. However, with advances in technology, now it is possible to collect learner data in large quantities, store it on the computer and analyse it automatically or semi-automatically using current available linguistic software.

> "Computer learner corpora are electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in the standardised and homogeneous way and documented as to their origin and provenance" (Granger, 2002, p. 7).

Sardinha (2004) indicates that learner corpora redefine the original concept of a corpus, which presupposed only native language. Therefore, the compilation of non-native speaker corpora enables the investigation of very pertinent research questions:

1. What are the features of the target language used with more frequency (overuse) or with less frequency (underuse) when compared to those of native speakers?
2. To what extent does the mother tongue influence learners' production in the target language, i.e. language transfer?
3. In which areas of the target language do learners tend to use strategies for avoidance, not exploiting the full potential of the target language?
4. In which areas of the target language do learners tend to demonstrate native-like and non-native-like performance?
5. What are the areas of the target language which learners from a specific country seem to require more pedagogical assistance so as to develop their production in the target language?

Sardinha (2004) explains that the notions of "overuse" and "underuse" are indicators of a reference standard deviation, which is usually the native language model. However, this deviation may also refer to the production of experienced native or non-native speakers, accepted by specific discursive or scientific communities. Therefore, the notions of "overuse" and "underuse" have a descriptive rather than prescriptive nature, allowing a better understanding of learners' interlanguage, as well as a diagnostic nature, offering data to teaching methodologies. "Overuse" and "underuse" oppose the notion of "error" (previously considered a negative deviation) as they are based on the principle that the learning of a second language necessarily involves the making of errors, not being an anomaly in acquisition. Learner corpora research sustains the inevitability of errors in the process of acquisition and describes such deviations in a neutral and substantiated way.

The above research questions presented by Sardinha (2004) were the starting point for the elaboration of the research questions of this study. Despite being broad in scope, they highlight the importance of corpus-based investigations in the description of learner production, in the specification of learners' strengths and weaknesses and in the comparison between learner and native speaker production. As spoken discourse is also an extremely broad area, research sub-questions were then further elaborated so as to focus more specifically on discourse, pragmatic and interactional phenomena which seem to be prototypical of the different facets of conversational competence and which are known to perform pragmatic functions.

Let us now describe the design principles when compiling corpora for research purposes.

3.1.2. Compiling and building corpora

Lee (2010) describes different generations of corpora. The first generation of "standard corpora" was compiled following the same sampling lines and set the parameters for corpora to come. The corpora contained 500 texts, of about 2,000 words each, sampled across a wide variety of written genres. "The Brown Corpus of written American English" is the most famous example of this generation as it was the very first corpus to ever be launched, in 1964. The second generation of corpora belongs to the Internet age. Consequently, the corpora are much larger. "The Bank of English" was the first mega corpus launched in 1991, covering

both spoken and written English and different regional varieties (British, American and other varieties). It was a continuation of the COBUILD corpus at the University of Birmingham, which had started earlier in 1980. "The British National Corpus" (BNC) was launched in 1995, presenting 100 million words, of which a tenth was composed of transcribed speech. The BNC is one of the most heavily used and researched corpora. The ten-million-word spoken component has both a demographically sampled component (mostly casual conversation) and a context-governed component (task-oriented) in a 6:4 ratio. The written component was mainly drawn from published sources.

Nowadays, there are already a large amount of corpora[29] available for use. Some corpora are free of charge, some have to be bought and some are not publicly available. Therefore, before deciding to build a corpus, it is a sensible idea to investigate existing corpora and whether they are suitable for specific research purposes. Learner corpora can be composed of written or oral texts, or a mix of both. In order to make a basic corpus, O'Keeffe, McCarthy and Carter (2007) explain that corpus programmes such as "Wordsmith Tools" (Scott, 1999) and "Monoconc Pro" (2002) enable computers to manipulate spoken or written texts with relative ease. The compilation of written texts is a relatively easy task as they may be scanned, typed or downloaded from the Internet or simply used in their existing electronic forms. The building of spoken corpora, on the other hand, is a far more time-consuming activity as it requires the audio-recording of live oral interactions and then their transcriptions. O'Keeffe, McCarthy and Carter (2007) indicate that one hour of recorded speech usually results in approximately between 12,000 to 15,000 words of data and may take around two days to transcribe, depending on the level of details to be included. For instance, a spoken corpus may be coded for different speaker turns, interruptions, speaker overlaps, truncated utterances, paralinguistic features such as giggling and yawing, external noise and prosodic features. The more details to be included, the longer the transcription will take. As a consequence, written corpora outnumber spoken corpora in quantity and size.

As previoulsy mentioned, corpora can be used for both quantitative and qualitative analysis research. Quantitative methods include examining the occurrence of lexical items in a specialised corpus. When results of occurrences are then compared with frequencies in other corpora, conclusions are drawn by qualitative means. O'Keeffe, McCarthy and Carter (2007)

[29] For a detailed description of available corpora, see O'Keeffe, McCarthy and Carter (2007, pp. 284-296).

advise that a selected corpus must best represent the language or the language variety and that the corpora themselves must be comparable. For instance, let us imagine that a teacher would like to investigate the overuse of the adjective "*nice*" in narratives by her/his EFL students. She/he would initially collect a soft copy of their compositions, store their work onto her/his computer and run a word count programme. The results of occurrence of "*nice*" would then be compared with other corpora such as British or American teenagers' narratives. If these results were compared with a corpus composed of Ernest Hemingway's novels, for instance, this would probably lead to flawed results.

When selecting, compiling or building a corpus to be used for analytical purposes, the question of corpus size arises. How big must a corpus be so that results are considered representative and reliable? The first thing to consider is the purpose and the nature of the search. If the aim is to examine a common word or expression, i.e. core vocabulary, in a much more specialised language register, a small corpus will suffice. However, if the search is intended to reveal the frequency of a less-frequent word such as "*maverick*", a much larger corpus will be needed. O'Keeffe, McCarthy and Carter (2007) add that the question of size really depends on whether the corpus is written or spoken and on what it is seeking to represent. Spoken corpora are considered large if they display over a million words whereas written corpora with below five million words are considered quite small. As for suitability, it is the design of the corpus rather than its size that is the determining factor.

Before building spoken corpora, O'Keeffe, McCarthy and Carter (2007) indicate that a corpus will initially require some design principle based on factors such as what researchers are intending to represent and how best this can be represented. Secondly, the mode of recording needs to be considered. Traditional analogue cassettes are cheap but difficult to store and cannot be easily computerised and aligned with transcriptions. Digital recording devices allow sound files to be stored with ease and to be aligned with transcriptions later on. In both cases, speakers should give their permission for the recording by signing off a consent form that authorises the use of the recorded material (transcript, voice or both) for research or pedagogical purposes. Thirdly, spoken data needs to be transcribed manually, which can be quite a challenging task as previously mentioned. Storing files as "plain texts" offers the maximum flexibility of use with different software suites. The level of detail to be included in the transcription will depend on the purpose of the research or of the corpus. For instance, if

the phenomenon "overlap" is not going to be investigated, it is unnecessary to spend time marking it on the transcription[30]. Once files have been transcribed, they will need to be organised in a way so that the source information can be traced. The authors advise researchers to retrieve information such as gender, age, number of speakers, place of birth, occupation, level of education, where the recording took place, relationship among speakers and to store it at the beginning of each transcript as an information "header" or in a separate database, where the information is logged with the file name. Lastly, the transcription will require careful checking with the original recording for accuracy.

Now let us describe corpus techniques that enable researchers to explore corpora for research purposes.

3.1.3. Corpus procedures

It is important to clarify that corpora in themselves do not inform researchers of anything as they are simply collections of electronic texts. Thus, researchers need to be familiar with corpus techniques and software in order to carry out linguistic investigations. The use of corpus software enables analytical investigations and comparisons of corpora data to be made. According to Evison (2010, p. 122), the first analytical steps usually involve two related processes: the production of frequency lists and the generation of concordances. These processes can be carried out by an increasing range of software, from commercial software such as "WordSmith Tools" (Scott, 1999), "Monoconc Pro" (2000) and "Word Sketch Engine" (Kilgarriff et al., 2004) to free software like "AntConc 3.4.1." (Anthony, 2014), which can be downloaded from the Internet. These two corpus-techniques are "built on the very basic foundation that electronic collections of texts can be searched very rapidly" (Evison, 2010, p. 122). The automatic generation of frequency lists can quickly produce a complete list of all the items in a corpus, from the most to the least frequent lexical items. The analysis of concordance begins with the investigation of a specific item; this search then displays all the examples of this item in context. It is the combination of these two basic corpus techniques with associated analysis (which may include both quantitative and qualitative methods) that provides researchers with insights on language in use.

[30] I followed the conventions for transcribing audio files suggested by O'Keeffe et al. (2007, p. 6) in the transcription of the interviews comprising the Brazilian learner corpus of this study, to be detailed in sub-section 3.3.3.

With regards to frequency lists, the software searches every item in that corpus so as to calculate how many "tokens" there are in total and how many different "types" comprise this total. To illustrate the distinction between "token" and "type", let us analyse the sentence initiating this paragraph: ("*With regards…total*"). When we add up the words, we can identify 32 tokens as "tokens" refer to words separated by spaces or punctuation, including the repetition of the same word. However, the very same sentence contains only 27 types as "types" refer to different forms (the words "*to*", "*in*", "*how*", "*many*" and "*total*", which are repeated, count only once). The software then outputs the final counts as a frequency list displayed in a rank order of frequency or in alphabetical order. In addition to the rank order and the raw frequency of each token, it is possible to see the percentage of tokens in the whole corpus that each frequency count represents.

According to Evison (2010), the analysis of alphabetical frequency lists when compared to rank of frequency lists raises two important issues. Firstly, it shows that corpus software counts tokens that may or may not be words in the traditional sense such as backchannels[31] or hesitation devices, i.e. "*uhuh*" and "*hum*". Most software will also count contractions as single items such as "*I'll*" and "*I'd*". Although corpus linguists usually consider contractions as single words as attested by corpus-based grammars, traditional grammars may consider them as two separate words. Secondly, there is the problem of lemmatisation[32]. O'Keeffe, McCarthy and Carter (2007) indicate that while words like "*look(s)*", "*looking*" and "*looked*" will be considered different words in frequency lists, these inflected forms may be brought together semi-automatically and treated as a lemma. However, the process of lemmatisation may group together words that the computer judges to look similar like "*man*" and "*mane*". The software may also fail to notice obvious similarities as in "*young/younger/youngest*". Thus, the authors suggest that taking lemmas as a benchmark may be problematic.

As for the usefulness of frequency lists, Evison (2010) states that these are precious sources for lexicographers, language syllabus and materials design as they provide researchers with a range of frequency information. They also form the basis of more complex statistical measures that express the strength of collocations, i.e. the likelihood of items co-occurring. It

[31] The quantitative analyses of native speaker corpora in chapter 6 showed that very few occurrences of the backchannel "*uhuh*" had actually been included in the transcriptions of spoken discourse.
[32] "Lemmatisation" refers to "matching any form of a word, with all its inflections, to a base form, usually the form that would be found in a dictionary entry" (Walter, 2010, p.432).

is also useful to compare the rank order of specific items in two or more corpora by looking at them side by side. This comparison allows conclusions to be drawn on the use of specific items in different genres (more intimate versus academic conversations) and on the nature of the genres themselves. However, a process of "normalisation" is required so as to compare frequency counts in corpora of different sizes. "Normalisation" involves "extrapolating raw frequencies from the different-sized corpora which are being compared so that they can be expressed by a common factor such as a thousand or a million words" Evison (2010, p. 126).

Furthermore, it is also possible to investigate the keyness of words. "Key words" are not necessarily the most frequent words in a corpus and are indentified as such by "statistical comparison of a "target" corpus with another, larger corpus, which is referred to as the "reference" or "benchmark" corpus" (Evison, 2010, p. 127). The "WordSmith Tools Keyword" programme automatically compares wordlists including items that are either significantly frequent (positive keywords) or infrequent (negative keywords). Although there are different ways of calculating statistical significance available (as previously described in footnote 28), Evison (2010) clarifies that a test of keyness is particularly useful for corpus data analysis, being based in a log-likelihood test rather than on the assumption that data have a normal distribution.

Another very popular corpus analytic technique is "concordance analysis", also known as "key word in context analysis" (Evison, 2010). A large number of websites already offer internet users the chance to search their corpora for specific words or phrases. The "COBUILD Concordance and Collocations Sampler" and the "Corpus-based Concordances", which partially comprise the "Compleat Lexical Tutor", are two examples of online concordancing programmes available for teachers and students. These websites enable users to search for concordances within their corpus but do not allow corpus files to be downloaded. The BNC can be both purchased and freely searched online. However, there are restrictions on the number of results to be displayed when doing online concordancing and only random samples may be available. Researchers may then carry out multiple searches for the same item to generate different random samples, which can be accumulated so as to validate results. The Michigan Corpus of Academic Spoken English (MICASE) does not limit the number of hits when concordancing but offers the possibility to limit the number of results when a very frequent item is being searched for.

For Evison (2010), the major advantage of concordancing as an analytical technique is that it enables a large number of examples of an item to be brought together in one place, in their original context. Concordancing is particularly useful for generating and testing hypotheses as a hypothesis can be generated based on observed patterns in a few lines and subsequently tested through further searches. However, Huston (2002) points out that when testing a hypothesis, it is essential to take into account the items that do not seem to support the hypothesis being tested, and if necessary, review the hypothesis rather than reject the items themselves.

A concordance programme searches an item within a corpus and displays the results of that search, i.e. the concordances or the concordance lines, on the screen. All the occurrences of the target item (or node word) are displayed, vertically centred, along with the present number of characters either side. When concordance lines are then sorted out, it is easier to identify regularities of occurrence. O'Keeffe, McCarthy and Carter (2007) state that most corpus software enables the number of words at either side of the node word to be adjusted so as to allow more of the text to be viewed. It also facilitates the sorting of the concordance lines so that the lexico-grammatical patterns occurring before and/or after the node word can be investigated. Evison (2010) highlights the versatility of concordance programmes as they have the potential to provide both insight into the typicality of item use and evidence of the most frequent meanings and the most frequent collocates, i.e. co-occuring items.

To sum up, in this section we have presented different views on the status of Corpus Linguistics. For Granger (2002), Corpus Linguistics is a linguistic methodology grounded on the use of electronic collections of naturally occurring texts. For Biber, Conrad and Reppen (1998), the corpus-based approach emphasises language in use by investigating how users exploit the resources of their language. Corpus-based investigations are empirical; they utilise a corpus as the basis for analysis; they make extensive use of computers for analysis; they depend on both quantitative and qualitative analytical techniques. For McEnery and Hardie (2012), Corpus Linguistics is an area that focuses upon a set of procedures and methods for studying language and has the potential to reorient one's approach to the study of language, to refine and redefine theories of language and to enable the use of theories of language that were previously difficult to be explored without corpus-linguist methods.

We have also addressed learner corpora, the compilation of corpora and corpus procedures. Learner corpora are electronic collections of authentic foreign language or second language texts assembled according to explicit design criteria for a particular foreign language or second language purpose (Granger, 2002, p. 7). When deciding to compile spoken corpora, O'Keeffe, McCarthy and Carter (2007) indicate that researchers need to take into account some design principle based on factors such as what they are intending to represent and how best this can be represented, the mode of recording and transcription conventions, as spoken data needs to be transcribed manually. In order to explore corpora for empirical studies, researchers rely on traditional corpus-based techniques such as the production of frequency lists and the generation of concordances lines and employ both qualitative and quantitive methods of analysis.

Now that we have described some key aspects of Corpus Linguistics, let us move on to discuss different approaches to the analysis of spoken discourse, which include corpus-based procedures.

3.2. INVESTIGATING SPOKEN DISCOURSE THROUGH COMBINED APPROACHES

As previously mentioned in section 1.1, the term "discourse" is used in different ways which reflect different perspectives. Within linguistic perspectives, discourse can refer to "language in use" (Brown & Yule, 1983), "language above the sentence or clause" (Stubbs, 1983) or specific types of language use or topics. In a sociocultural view, Critical Discourse Analysis (Fairclough, 1995) investigates the connections between language, power and ideology. Thus, the above views of discourse point to different foci of studies encompassing specific phenomena, which may or may not lend themselves to corpus-based investigations. Considering spoken discourse in particular, interactional phenomena like pair-adjaceny and turn-taking are traditionally investigated within CA domains. Likewise, pragmatic phenomena like conversational implicatures do not seem to fit into corpus-based investigations. However, there are some discourse, pragmatic and interactional phenomena that are triggered by individual lexical items (including pragmatic markers), lexical bundles, clause patterns and formulae, and thus, may be retrieved by means of corpus searches.

This section describes some possible interfaces of corpus-based investigations with discourse, pragmatic and interactional studies, in which approaches may be combined.

3.2.1. Corpus Pragmatics

Broadly speaking, Jucker (2013, p. 1) relates Corpus Pragmatics to the "studies of language use that employ large, computer-readable collections of language". Such studies are quite recent and differ from philosophical approaches to pragmatics as they are placed within empirical data-based pragmatics. Jucker (2013) explains that pragmaticists have taken some time to realise the potential of large corpora for pragmatic investigations since they have traditionally worked with philosophical methods for the analyses of pragmatic phenomena or focused on micro-analyses of conversational exchanges. Likewise, corpus linguists have also taken some time to develop an interest in pragmatic phenomena as they would initially adapt their research questions to corpus linguistics methodology by basically focusing on easily retrievable structures.

The core of pragmatic investigations lies in the use and the functions of language. "Form-to-function mapping" refers to investigations of particular linguistic entities and their functions, for instance, the uses of a specific discourse marker. Conversely, "function-to-form mapping" refers to investigations of specific functions and the range of linguistic devices employed to perform them, for example, the realisations of a specific speech act. Jucker (2013) points out that form-to-function research questions lend themselves more easily to corpus linguistics investigations since it is relatively easy to translate them into a search string. Function-to-form research questions, on the other hand, are far more challenging as they do not automatically turn into search strings. In addition, since pragmaticists tend to be more often interested in the spoken variety, the further difficulty of having to transcribe spoken language may also account for a relative small number of studies in Corpus Pragmatics.

Jucker (2013) points out that Aijmer's (2002, 2008, 2009) studies are very explicitly based on corpus, and thus, illustrate Corpus Pragmatics form-to-function investigations. By conducting searches in two different corpora, London - Lund (a corpus of spoken British English from the 1960s comprising half a million words) and Lancaster-Oslo/Bergen (a one-million word

corpus of written British English), Aijmer (2002)[33] uses corpus evidence to provide a sufficiently large basis of examples of the discourse particle *"now"*, the interjections *"oh"* and *"ah"*, evidential or heading particles like *"sort of"* and *"kind of"* , the expectation marker*"actually"* and the interpersonal particle *"just"*. While the focus of analysis remains on the development of functional profiles of the above elements, statistical results are also included such as the frequency of individual elements in text types or various collocational patterns. Following from this, Aijmer (2008) continues her corpus-based investigations by using a one-hundred million word corpus, the British National Corpus, while focusing on evaluative markers such as *"amazingly"*, *"surprisingly"* and *"astonishingly"*. Aijmer (2008) includes data from the written component of BNC (ninety million words) as the frequency of the above evaluative markers is much higher when compared to that of the spoken component of the corpus (10 million words). Aijmer (2009) extends her analysis of the interjections *"oh"* and *"ah"* from the London- Lund Corpus to the half-million word Bergen Corpus of London Teenage Language (COLT), which features conversations among teenagers at London-based schools.

Jucker (2013) also describes Corpus Pragmatics investigations combined with Sociolinguistics. Andersen (2001) uses COLT to investigate the variation of the invariant tags *"innit"* and *"is it"* and the pragmatic marker *"like"*. Tottie (2011) assesses the frequency of the sociolinguistic markers *"uh"* and *"um"* in British English and American English corpora. Her results indicate that these markers are more frequently produced by men, by older speakers, by speakers from a higher socio-economic background and by speakers with a higher education. Culpeper and Kytö (2010)[34] carry out a diachronic investigation of pragmatic noise in the Corpus of English Dialogues 1560-1760. Culpeper and Kytö (2010, p. 199) define pragmatic noise as elements which "do not have related words which are homonyms in other word classes, do not participate in traditional sentence construction, are morphologically simple and have less arbitrary meanings compared with most words". Culpeper and Kytö (2010) trace elements such as "*ah*", "*ay*", "*alas*", "*ha*", "*oh*", "*um*", "*huh*", and "*hum*", analyse their functional profiles and attempt to understand the pragmatic effects these elements must have had over time. They also focus on the variability of these elements and provide a large range of frequency statistical data.

[33] Aijmer's (2002) findings on the pragmatic functions of "*actually*" and "*just*" are detailed in chapters 4 and 5, respectively.
[34] Culpeper and Kytö's (2010) summary of spoken conversation features was detailed in chapter 1.1.1.

As for function-to-form investigations, Jucker (2013, p. 7) explains that many pragmatic phenomena like speech acts are defined through their function. He presents three ways to retrieve specific speech acts from large corpora: focusing on specific illocutionary force indicating devices, focusing on syntactic/semantic patterns that have been shown to be typical for specific speech acts and retrieving relevant speech act verbs used performatively to carry out a specific speech act or descriptively. Aijmer (1996) analyses the speech acts of thanking, apologising, requesting and offering which, in English and in other languages, can be carried out in highly conventionalised forms, i.e. illocutionary force indicating devices, such as "*thank you/thanks*", "*sorry*" and "*please*". Typical patterns, on the other hand, go beyond the category of IFIDs. Manes and Wolfson (1981) claim that compliments in American English are highly conventionalised and that the pattern "NP {is/looks} really ADJ" accounts for more than half of their examples. Although such search strings can be used to retrieve specific speech acts, they are problematic as they depend on a list of known patterns by the researcher. Transforming the notation of a semantic/syntactic pattern into a precise search string is not a straightforward task either. Lastly, Jucker and Taavitsain (2014) investigate compliments in the history of American English. They search for the word *"compliment"* in the 400-million-word Corpus of Historical American English (COHA) and in the 425-million-word Corpus of Contemporary English (COCA). Their results indicate that the word *"compliment"* is more often used discursively or descriptively than as a performative verb.

To conclude, in this sub-section we have presented Corpus Pragmatics and illustrated form-to-function and function-to-form corpus-based investigations of pragmatic and sociolinguistic phenomena. Corpus Pragmatics refers to the "studies of language use that employ large, computer-readable collections of language" (Jucker, 2013, p. 1). Form-to-function studies (Aijmer 2002, 2008, 2009, Culpeper & Kytö, 2010) focus on particular linguistic entities and their functions. They lend themselves more easily to corpus linguistics investigations since it is relatively easy to translate them into a search string. Function-to-form studies (Aijmer, 1996, Jucker & Taavitsain, 2014) examine specific functions and the range of linguistic devices employed to perform them. Although function-to-form investigations are far more challenging as they do not automatically turn into search strings, they may be retrieved through specific syntactic/semantic patterns or protopypical lexical items known to perform such functions.

In the next sub-section, I present a conjoint approach (O'Keeffe & Walsh, 2012) to the study of spoken discourse, which relies on both Corpus Linguistics and Conversation Analysis.

3.2.2. Corpus Linguistics and Conversation Analysis as a conjoint approach

In addition to a combined approach to pragmatic and sociolinguistic investigations, Corpus Linguistics can also provide very relevant data to researchers who wish to investigate interactional features. O'Keeffe and Walsh (2012) indicate that Conversational Analysis and Corpus Linguistics have different origins and focus on different research objects. While CA analysts may use a very small portion of naturally occurring data, they aim at revealing how interactants co-construct meanings, repair breakdowns and orient to each other to maintain a sequential order to achieve a communicative purpose. Corpus Linguistics researchers, on the other hand, also rely on naturally occurring data (usually much larger amounts of data) but they investigate specific linguistic features in terms of word frequency, concordances, multi-word units and keyness through quantitative analyses.

Nonetheless, O'Keeffe and Walsh (2012) clarify that both CA and Corpus Linguistics share methodological similarities and can be used as complementary approaches: they use empirical, naturally occurring data; they refer to baseline comparisons with other types of interactions (sequential order and reference corpora, respectively); they investigate language in its social context; analyses are carried out at turn level for a better understanding of context. To illustrate this, O'Keeffe and Walsh (2012) analyse discourse in the context of longer stretches of text, combining corpus techniques with CA. Their interest lies in the investigation of features of spoken interactions, such as adjacency pairs, in higher education small group teaching sessions. The term "CLCA" refers to this conjoint approach, which, in their view, offers a more "up-close" description of spoken interactions in context than would have been normally provided if each approach had been employed on its own. By following this conjoint approach, O'Keeffe and Walsh (2012, p. 161) are able to "gain powerful insights into the ways interactants establish understandings and observe how words, utterances and text combine in the co-construction of meaning".

For O'Keeffe and Walsh (2012), Corpus Linguistics has increasingly been applied to different domains and contexts, traditionally associated with other methodologies. In such cases,

Corpus Linguistics has successfully been used alongside other approaches like CA, discourse analysis, critical discourse analysis and pragmatics. In such studies, Corpus Linguistics has been applied to achieve a particular goal rather than to describe the language features of a given corpus. The authors draw a methodological distinction between descriptive corpus research and applied corpus research. In the former case, the corpus is an end in itself. In other words, researchers investigate a corpus so as to analyse the use of language and to detail patterns in a particular genre. In the latter one, the corpus is a means to an end. Researchers look beyond the corpus for both its research questions and its analysis.

Another merit of the CLCA approach is its appropriateness for research into spoken discourse. O'Keeffe and Walsh (2012) explain that in the past, written corpora were utilised to investigate lexical patterning rather than discourse context. Larger corpora were used in order to attain the greatest breadth of semantic coverage for lexical items. Conversely, there is a current trend to use smaller, specialised and highly context specific corpora, which are frequently based on spoken interactions. O'Keeffe and Walsh (2012) suggest that such corpora are particularly suitable for a conjoint approach (such as Corpus Linguistics with CA), which allows the investigation of other discourse features. In smaller corpora (between 50,000 to one million words), the context is more clearly defined, enabling the investigation of discourse features that operate beyond sentence level. The person who has recorded and transcribed the data is often the same person who is investigating interactional discourse, and thus, has more detailed information about the interactional and social context, including information about speakers, their roles and relationships and about the purpose of their interactions.

In this sub-section we have presented the conjoint approach "CLCA" (O'Keeffe & Walsh, 2012) to the study of spoken discourse. While Conversational Analysis and Corpus Linguistics have different origins and focus on different research objects, they share methodological similarities and can be used as complementary approaches. The term "CLCA" refers to this conjoint approach, which offers a more "up-close" description of spoken interactions in context than would have been normally provided if each approach had been employed on its own. Another merit of the CLCA approach is its appropriateness for research into spoken discourse. There is a current trend to use smaller, specialised and highly context specific corpora, which are frequently based on spoken interactions. Thus, such corpora are

particularly suitable for a conjoint approach that allows the investigation of other discourse features.

Now that I have justified the choice of Corpus Linguistics in combined approaches to the investigation of spoken discourse, including the investigation of discourse, pragmatic and interactional phenomena, let me focus on the empirical study of this thesis by describing its aims, subjects, instruments and corpora and the corpus procedures employed.

3.3. THE EMPIRICAL STUDY

This thesis includes an empirical study, whose purpose is to investigate some of the key discourse, pragmatic and interactional phenomena comprising conversational competence (proposed in section 2.3.2) in the oral production of twenty Brazilian learners of English at CEFR B1. As previously mentioned in section 1.1, this thesis is grounded on a functional approach to discourse analysis (Brown & Yule, 1983), which focuses on the analysis of language in use and describes linguistic forms along with their functions. In order to accomplish the above objective, research questions were initially elaborated bearing in mind functions that are likely to be performed in the management of discourse, in the negotitation of illocutionary meaning and in conversational practices. Then, research sub-questions were further elaborated so as to narrow the scope of investigation to prototypical discourse, pragmatic and interational phenomena known to perform such pragmatic functions and which may be activated when learners engage in conversations. In addition, the selected phenomena “discourse marking adverbs”, “adverbial hedges” and “minimal response tokens” are sub-types of pragmatic markers that lend themselves to corpus-investigations since they consist of lexical items and thus, can be more easily retrieved.

Chapters 4, 5 and 6 address the research sub-questions of this study and include justifications for choices of specific lexical items under investigation, based on corpus evidence and frequency. Let us now focus on the study itself by detailing its subjects, instruments, corpora and the procedures employed in the empirical investigations.

3.3.1 Subjects

The present empirical study included the participation of twenty Brazilian learners of English over a one and a half year period. Data were collected between the second semester of 2012 and the second semester of 2013. All subjects attended a general English course at a language institute in the South of Brazil and their age ranged from 20 to 40 years old. Before joining the target course, new students at the institution had taken a placement test so as to ensure they were at the same linguistic level of competence as the other regular students, in this case, at CEFR B1 +. Learners attended two and a half hour-lessons once a week, which consisted of skills development, language work and communicative activities. I worked as their class teacher in the semester in which the interviews were recorded. 2012 students followed the course book "New Framework – level 4" (Jeffery et al., 2009) whereas 2013 students followed the course book "World Wise 1" (Cowley, 2012), which replaced the Framework series.

Most learners had already visited English speaking countries for tourism and wished to study English further for communication purposes. During the semester in which I worked as their teacher, there was no specific pedagogical intervention related to the aims of this investigation. I basically implemented the lesson plans provided by the institution, based on the approach "Communicative Language Teaching" [35], adapting lessons and activities according to students' profiles, interests and needs.

Now that I have described the subjects of this study, let me detail the instruments used to collect, analyse and compare data.

3.3.2 Instruments

The present study relied on three instruments for different purposes, namely the speaking test of "Cambridge English First" for data collection, the "Corpus Query Processor" web server for access to native speaker corpora and for corpora analyses and comparisons, and the free software "AntConc 3.4.1." (Anthony, 2014) for learner corpora analyses.

[35] Communicative Language Teaching encourages students to communicate rather than just practise certain language items. It teaches language for communication, e.g. language functions (Harmer, 2012, p. 262).

Data collection:

Before starting the process of data collection, the empirical project of this thesis was submitted to CEP-PUCRS (Comite de Ética em Pesquisa da Pontifícia Universidade Católica do Rio Grande do Sul) for approval since it included the participation of human beings. Once the approval was granted, subjects were informed of the project aims and procedures and then signed a consent form[36]. In order to collect samples of learners' spoken production, I selected a public version[37] of the speaking test of the examination "Cambridge English First", also known as "FCE" (University of Cambridge ESOL Examinations, 2007, pp. 79-81).

FCE is aimed at learners of English at CEFR B2. In the ESOL Common Scale for Speaking, successful candidates at CEFR B2 show a generally effective command of the spoken language (University of Cambridge ESOL Examinations, 2007, p. 87): they are able to handle communication in familiar situations; they can organise extended discourse but occasionally produce utterances that lack coherence and some inaccuracies and inappropriate usage occur; they maintain a flow of language, although hesitation may occur while searching for language resources; although pronunciation is easily understood, L1 features may be intrusive. FCE candidates participate in a paired face-to-face format test (two candidates and two examiners), which gives them the opportunity to demonstrate their spoken abilities in a controlled but friendly environment. While one examiner acts as the interlocutor conducting the oral interview, the other examiner acts as the assessor grading candidates' oral production according to analytical criteria.

The FCE speaking test has scripted rubrics, takes 14 minutes and consists of four parts. In the selected version of FCE[38], parts one and two take 3 minutes each whereas parts three and four take 4 minutes each. In part one, the interlocutor asks individual candidates personal questions related to their lives, studies, work, leisure time, interests and future plans. Candidates are able to show their ability to use general social and interactional language and to give basic personal information about themselves. In part two, each candidate receives two coloured photographs and talks about them for one minute without interruption (long turn), comparing

[36] For the consent form, see appendix A.
[37] For the selected version of the FCE speaking test, see attachment A.
[38] A revised version of FCE will be launched in the first semester of 2015.This new version includes minor changes to the time of speaking test parts.

the pictures based on a spoken rubric given by the interlocutor. Next, the listening candidate answers a short question (about 20 seconds) presented by the interlocutor on the same topic. The same procedure is then repeated with the second candidate. Candidates are given the opportunity to show their ability to produce an extended piece of discourse.

In part three, candidates engage in a collaborative task for three minutes. They are given oral instructions and provided with visual stimulus, usually several photographs or pieces of artwork. Candidates are expected to discuss each visual prompt, expressing and justifying their opinions, evaluating and speculating so as to reach a negotiated decision towards the end of the task. The interlocutor does not interfere in this part. Candidates are given the opportunity to demonstrate their ability to engage in a discussion and to work towards a negotiated outcome. Lastly, in part four, the interlocutor directs the interaction by asking further questions that encourage candidates to broaden the scope of the topic discussed in part three. This final part gives candidates the opportunity to show their ability to discuss the topic of the collaborative task in more depth.

For the purposes of this study, I decided to use the FCE speaking test as an instrument to collect samples of learners' oral production for three main reasons. Firstly, as a formal assessment instrument, the speaking test is designed and validated[39] to give candidates the opportunity to produce enough samples of language so as to have their speaking ability assessed in four analytical criteria, namely "grammar and vocabulary", "discourse management", "pronunciation" and "interactive communication". All four analytical criteria are described within three bands of performance that detail specific abilities. While "grammar and vocabulary" and "pronunciation" are traditionally viewed as code components within competence frameworks (Canale & Swain, 1980, Canale, 1983, Bachman, 1990), "discourse management" and "interactive communication" reflect use components. In a broad view, discourse management and interactive communication are described as follows (University of Cambridge ESOL Examinations, 2007, p. 86):

a) Discourse management: the candidate's ability to link utterances together to form coherent speech, without undue hesitation; the utterances should be arranged logically to develop the themes or arguments required by the task;

[39] For context validity, see Galaczi and ffrench (2011, pp. 112-170).

b) Interactive communication: the candidate's ability to take an active part in the development of the discourse; this requires the ability to participate in the range of interactive situations in the test and to develop discussions on a range of topics by initiating and responding appropriately. This also refers to the deployment of strategies to maintain interaction at an appropriate level throughout the test so that tasks can be fulfilled.

The performance descriptors of the two criteria above list elements that are also accounted for in the proposed model for conversational competence. The management of interaction comprises the notions of cohesion and coherence, cohesive devices, discourse markers while the deployment of conversational practices includes turn-taking, initiating and responding. Thus, phenomena inherent in these notions are likely to be activated when students carry out the speaking test tasks. However, a cautionary note to add is that the model for conversational competence proposed in section 2.3.2 also differs significantly from the above analytical criteria and their performance descriptors. It is intended to describe the role of key discourse, pragmatic and interactional phenomena in the three interrelated facets of conversational competence, within a pragmatic domain, rather than to provide performance descriptors for assessment purposes. It is based on the idea that meaning is co-constructed, and thus, it includes a third use component, namely the negotiation of illocutionary meaning, accounting for both production and comprehesion dimensions.

Secondly, the FCE speaking test is delivered in a paired format. Galaczi (2008, p. 91) explains that the construct of speaking ability underlying Cambridge English speaking tests draws on theoretical advances such as theories of communicative language ability (Bachman, 1990, Bachman & Palmer, 1996) and pedagogical developments. These theories of language ability include a conversation management component within specific competences (namely textual competence as previously discussed in chapter 2.2.). One of the advantages of a paired format is that it gives candidates the opportunity to display their conversation management ability. Galaczi (2008) clarifies that although the test as a whole is controlled by the interlocutor, there are parts that are controlled by the candidates, and thus, provide opportunities for peer-peer discourse.

Thirdly, Brazilian undergraduate students applying for the "Science without Borders" programme require a successful performance at specific CEFR levels, ranging from B2 to C1,

which is dependent on the demands of the target country institution and on the examination taken per se. Learners' spoken performance in tasks whose level of linguistic difficulty is slightly higher (CEFR B2) than that of their current level (CEFR B1 +) is likely to reveal the most obvious challenges faced by them in terms of moving beyond the intermediate plateau and it may also highlight conversational phenomena they already produce. Learners at the intermediate plateau usually have mixed linguistic abilities and may display some of the following characteristics (Richards, 2008, pp. 2-3): there is a gap between receptive and productive competence; there are persistent fossilised language errors; fluency may have progressed at the expense of complexity; the learner has a limited vocabulary range; language production may be adequate but often lacks the characteristics of natural speech. Thus, the findings of this study are intended to shed some light on the intermediate plateau phenomenon, providing teachers with an authentic sample of Brazilian learners' strengths and weaknesses in terms of spoken production at CEFR B1. The pedagogical recommendations proposed may be of some value for teachers whose students display similar difficulties.

Corpora analyses and comparisons:

During my PhD studies, I was awarded a scholarship from the Capes Foundation to conduct part of my doctoral research as a Visiting Research Student at Lancaster University, UK. As a graduate student at the Department of Linguistics and English Language, I was given access to the "Corpus Query Processor" web server (CQPweb so forth) and large and specialised corpora. CQPweb (https://cqpweb.lancs.ac.uk/) is "a new web-based analysis system, intended to address the conflicting requirements for usability and power in corpus analysis software" (Hardie, 2012, p. 380). CQPweb's most innovative feature is its flexibility, since its more generalised data model makes it compatible with any corpus (Hardie, 2012, p. 380).

A number of corpora are available on CQPweb server covering present-day English, historical English, learner English, Arabic, European languages and South and East-Asian languages. Users have access to specific corpora depending on their affiliation and profile. For my analyses and comparions, I used two of the available corpora: the spoken component of the "BNC Sampler" and a sub-corpus of the "Diachronic Corpus of Present-Day Spoken English". Both corpora will be described in the next section. The analysis options available in CQPweb include concordancing, collocations, distribution tables, frequency lists and key

words. In addition, the server offers 27 practical video tutorials, which teach users how to carry out analyses using the tools and the corpora available.

Nonetheless, CQPweb does not allow users to upload their own corpora. Thus, in order to conduct the analyses in the Brazilian learner corpus, I used the free software "AntConc 3.4.1." (Anthony, 2014). AntConc (http://www.laurenceanthony.net/antconc_index.html) is "a freeware, multiplatform tool for carrying out corpus linguistics research and data-driven learning" (Anthony, 2014, introduction). It runs on computers running Microsoft Windows (tested on Win 98/Me/2000/NT, XP, Vista, Win 7), Macintosh OS X (tested on 10.4.x, 10.5.x, 10.6.x) and Linux (tested on Ubuntu 10, Linux Mint). Users initially need to download the software to their computers. In order to access a compiled corpus using this software, corpus files should have been saved as plain text files ".txt". For searches, users open the programme, select specific files and then carry out analyses using a number of corpus tools: Concordance, File View, Clusters/N-Grams, Collocates, Word List and Key Word List.

Word List, Concordance and File View were the tools used for the analyses of the Brazilian learner corpus. The word list tool counts all the words in the corpus and presents them in an ordered list, which displays rank and raw frequencies. The concordance tool displays search results in a "KWIC" (Key Word in Context) format, which allows users to see how words and phrases are commonly used in the corpus. It is possible to limit the number of text characters to be outputted on either side of the search term, using the increase and decrease buttons on the right of the button bar under the "Search Window Size" title. The KWIC Sort option rearranges the concordance lines at three different levels: O (the search word), 1L, 2L… (words to the left of the target word) and 1R, 2R… (words to the right of the target word). Results are then shown in different colours, making patterns more visible and salient. The file view tool displays the raw texts of individual files so that results generated in the other tools can be investigated in more detail.

Now that I have described the instruments used for data collection, analyses and comparisons, let me detail the corpora that formed the core of the empiral investigations of this study.

3.3.3. Corpora

This study comprised both quantitative and qualitative analyses of empirical data retrieved from native speaker and learner corpora. A small specialised corpus of Brazilian learners´ oral production at CEFR B1 had been previously built for the purposes of this thesis. Spoken sub-corpora derived from "The BNC Sampler" and "The Diachronic Corpus of Present-Day Spoken English" were selected as native speaker benchmark corpora. As the subjects of this investigation had been exposed to textbooks and classroom materials based mostly on the British variety, it felt methodologically adequate to focus on British English native speaker corpora available on CQPweb for comparisons.

The BNC Sampler is a sub-set of the 100 million-word British National Corpus (BNC). BNC was produced as the result of a collaboration between dictionary publishers Oxford University Press, Longman and Chambers, research centres at Oxford and Lancaster Universities and the British Library. BNC was first launched in 1994 and then further revised, enhanced and reformulated in two subsequent editions. The written component (90%) displays extracts from newspapers, periodicals, journals, academic books, popular fiction, published and unpublished letters, memoranda and school and university essays. The spoken part of BNC (10%) includes a large amount of unscripted informal conversation together with spoken language in different contexts such as formal business and government meetings, radio shows and phone-ins.

The BNC Sampler is also a two-part corpus (written and spoken data), which was originally compiled to mirror the composition of the full BNC as far as possible. The BNC Sampler was initially used for a tagging enhancement project at Lancaster University. It was annotated with a more detailed set of part-of-speech tags than the BNC, manually checked and post-edited. It contains 2,304,126 words (52,125 types) spread in 184 texts. The spoken sub-corpus contains 1,140,055 words distributed in 98 texts. With regards to spoken interaction type, 80 texts display dialogues while 18 texts portray monologic discourse. This study used the spoken sub-corpus of the BNC Sampler as one of its benchmark corpora. However, as this sub-corpus includes samples of spoken language in different genres, a more specialised corpus of spoken conversation was also compiled from another spoken corpus.

The Diacronic Corpus of Present-day Spoken English (DCPSE so forth) was constructed at the Survey of English Usage, University College London. It consists of spoken data collected from the spoken section of the British International Corpus of English (ICE) and from the London-Lund Corpus of Spoken English (LLC). It contains 1,227,063 words spread over 280 texts. It portrays relatively contemporary spoken English covering a period of a quarter of a century, from the 1960s and early 1990s. Since CQPweb enables sub-corpora to be created, I used DCPSE corpus metadata information as a filter in the compilation of a specialised corpus named as "face-to-face conversations from the 1990s". This spoken sub-corpus contains 271,360 words distributed in 110 texts and was also used as a benchmark corpus.

As for the Brazilian learner corpus, I built a small corpus comprising 10 manually transcribed FCE interviews. At the stage of data collection, students performed the FCE tasks in a paired format (two students). I acted as the interlocutor delivering the interviews, which were recorded with a digital recording device. No assessor was present since the aim of conducting the interviews was to collect data. I basically followed the rubrics and the format of the FCE speaking test but allowed my students to speak for more than 14 minutes as my objective was to collect natural and spontaneous contributions rather than to reproduce exam conditions. At a later stage, I manually transcribed 2 hours 40 minutes and 14 seconds of oral speech, following the conventions for transcribing audio files suggested by O'Keeffe, McCarthy and Carter (2007, p. 6):

a) <S1> to <S21>

These marked the different speakers in the order they appeared in the numbered interviews. <S1> was the interlocutor while <S2> to <S21> were the subjects of this study. The interlocutor's contributions were not considered in the analyses.

b) +

This symbol marked interruptions from where they occurred and from where the utterances were resumed.

c) =

This symbol marked unfinished or truncated words.

d) <?>

This symbol marked unintelligible utterances.

e) <SE> … </SE>

The above pattern was used to mark extra-linguistic information: laughing, giggling, coughing, overlapping talk, unfilled pauses, voice fading, etc.

In addition, filled pauses and backchannels were also included in the transcriptions. For the purposes of this thesis, no phonetic symbols were indicated. It is also important to stress that learners' oral production was reproduced as originally uttered by them. No attempt was made to correct their grammatical, lexical, phonological and pragmatic mistakes.

Once the transcription of the 10 FCE interviews was finalised, individual files were carefully revised and then saved as plain text files (.txt). The final outcome was a small specialised corpus of 23,803[40] word tokens, displaying Brazilian learners' oral production at CEFR B1 +. For comparative purposes, relative frequencies derived from the Brazilian learner corpus and the two benchmark corpora have been normalised to 10,000 words due to the differences in corpus size.

In addition to the above three corpora, other learner and native speaker corpora were referred to in specific chapter sub-sections. The Louvain International Database of Spoken English Interlanguage is a corpus of informal interviews with upper intermediate and advanced learners of English from 11 mother tongue backgrounds: Bulgarian, Chinese, Dutch, French, German, Greek, Italian, Japanese, Polish, Spanish and Swedish. It includes over one million words, of which almost 800,000 were produced by learners. It consists of 554 interviews, which correspond to 130 hours of recording. Each interview consists of three tasks: a warming-up activity in which learners were asked to talk about one of three set topics for a few minutes, a free informal discussion conceived as the main part of the interview and a picture description. Data from the Swedish component of LINDSEI (Aijmer, 2011) was referred to in chapter 4.1.1 and compared to the findings in its native speaker counterpart "The Louvain Corpus of Native English Conversation". LOCNEC is a 170,533 word corpus that replicates LINDSEI tasks but has young native speakers of British English as interviewees instead of learners. Data retrieved from the Spanish component of LINDSEI[41] was also referred to for illustrative purposes in other chapter sub-sections.

[40] 23,803 word tokens as displayed in the freeware "AntConc 3.4.1." (Anthony, 2014).

[41] I was given access to LINDSEI at the Centre for Corpus Approaches to Social Science (CASS) at Lancaster University. The available version does not include Brazilian learners of English.

The Bergen Corpus of London Teenage Language (COLT) contains half a million words and it is a constituent of the BNC spoken corpus. It displays the speech of 13 to 17-year old teenagers from different boroughs of London. COLT data (Paradis, 2003) was included in section 4.1.2. The London-Lund Corpus of Spoken English (LLC) displays British English and contains 500,000 words spread in 100 texts. It has been prosodically annotated. LLC data (Aijmer, 2002) was included in sections 4.1.3 and 5.1.2.

The Cambridge and Nottingham Corpus of Discourse in English (CANCODE) is a five-million word corpus displaying British spoken discourse while the Limerick Corpus of Irish English (LCIE) is a one-million word corpus of contemporary spoken Irish English. Both corpora were referred to in chapter 6 (O'Keeffe & Adolphs, 2008, McCarthy 2002).The Corpus of Contemporary American English is the largest freely-available corpus of American English (http://corpus.byu.edu/coca/). It contains 450 million words and it is equally divided among spoken, fiction, popular magazines, newspapers and academic texts. COCA data was included in section 6.1.2 due to insufficient transcriptions of minimal response tokens in both British corpora.

Now that I have described the corpora used in both quantitative and qualitative analyses of this study, let me finalise this section by detailing some procedures related to the empirical investigations.

3.3.4. Procedures for empirical investigations

Chapters 4, 5 and 6 address the research sub-questions of this thesis. Chapter 4 aims at specifying the most common discourse marking adverbs used to mediate segments of discourse in conversations by Brazilian learners at CEFR B1. It initially presents views on "discourse marking adverbs", "stance adverbs" and comparisons of quantitave data from the the Brazilian learner corpus and the benchmark corpora. Once three adverbs are indicated for investigation, namely "*well*", "*really*" and "*actually*", it moves on to introduce qualitative analyses following a form-to-function approach.

Chapter 5 focuses on the most common explicit and implicit adverbial hedges used to mitigate representative speech acts by Brazilian learners at CEFR B1. It briefly revisits explicit and

implicit pragmatic force modifiers and then presents a number of forms traditionally associated with mitigation, based on corpus-evidence. The rank position of likely candidates for hedges in the Brazilian corpus serves as a starting point for the selection of forms to be investigated. Once forms are qualified and specified, it moves on to describe the degree of explicitness and implicitness of the hedges "*probably*", "*maybe*" and "*just*" as well as their pragmatic functions through qualitative analyses.

Chapter 6 describes the most common minimal response tokens used to express good listenership by Brazilian learners at CEFR B1. It initially introduces forms and functions of response tokens, drawing a distinction between minimal and non-minimal forms. In order to avoid some possible overlap with discourse markers, since these may be used as non-minimal response tokens, minimal forms are then selected for investigation. Based on raw frequencies in the Brazilian learner corpus, the minimal forms "*yeah*" and "*uhuh*" are chosen for qualitative analyses, following a form-to-function approach.

In order to signal specific interviews (1 to 10) and FCE test parts (1 to 4), the following code was used in all qualitative analyses of the Brazilian learner corpus and placed at the end of the transcribed extracts, for instance (FCE P1 - B1 BR-03). This code is to be interpreted as follows:
P1: Part 1
B1: CEFR level B1
BR: Brazilian learners of English
03: interview number 3 (out of 10)

Last but not least, each of the above chapters discusses its findings in the light of specific research sub-questions. These results are intended to provide evidence of pragmatic markers-used to mediate segments of discourse (discourse marking adverbs), to mitigate representative speech acts (hedges) and to express good listenership (minimal response tokens) - which seem to pose a challenge to Brazilian learners of English at CEFR B1. Each chapter then ends with some pedagogical recommendations.

To sum up, in this section I have introduced the empirical study of this thesis and described its subjects, instruments, corpora and overall procedures employed in the corpus-based

investigations. This study included the participation of 20 Brazilian learners of English at CEFR B1+ at a language institute in the South of Brazil. The FCE speaking test was used in its paired format as a means to collect learner oral production. The 10 digitally recorded FCE interviews were manually transcribed following the conventions proposed by O'Keeffe, McCarthy and Carter (2007). The transcriptions included backchannels and paralinguistic information such as overlapping talk, unfilled pauses and laughter. The benchmark corpora investigations relied on the CQPweb server while the Brazilian learner corpus was analysed using the free software AntConc 3.4.1. (Anthony, 2014). The benchmark corpora consisted of two British English native speaker sub-corpora, namely the spoken component of the "BNC Sampler" and the face-to-face conversations from the 1990s, compiled from the "Diacronic Corpus of Present-day Spoken English". Other native speaker corpora and learner sub-corpora (LINDSEI) were referred to for specific purposes in chapter sub-sections. The main procedures employed in the empirical investigations will be further addressed in the next 3 chapters.

SUMMARY

In this chapter, I have focused on the methodology and the procedures employed in the empirical investigations of this study. Corpus Linguistics is an area that focuses upon a set of procedures and methods for studying language (McEnery & Hardie, 2012). It has the potential to reorient one's approach to the study of language, to refine and redefine theories of language and to enable the use of theories of language that were previously difficult to be explored without corpus-linguist methods. Corpus-based analyses are empirical, utilise a corpus as their basis, make extensive use of computers and depend on both quantitative and qualitative techniques (Biber et al., 1998). Qualitative research uses data only as a basis for identifying and describing aspects of usage in the language whereas in quantitative research, features are classified, counted and more complex statistical models are built in an attempt to explain observed phenomena (McEnery & Wilson, 1996).

The investigation of learner corpora informs both second language acquisition (SLA) research and foreign language teaching research. Learner corpora are electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose (Granger, 2002, p. 7). The compilation of non-native speaker corpora

enables the investigation of pertinent research questions (Sardinha, 2004). The use of corpus software allows analytical investigations and comparisons of corpora data to be made (Evison, 2010). The first analytical steps usually involve two related processes, namely the production of frequency lists and the generation of concordances, which can be carried out by an increasing range of software. The automatic generation of frequency lists can quickly produce a complete list of all the items in a corpus. The analysis of concordance begins with the investigation of a specific item. It is the combination of these two basic corpus techniques with associated analysis (which may include both quantitative and qualitative methods) that provides researchers with insights on language in use (Evison, 2010).

Following from this, combined approaches to the investigation of spoken discourse were presented. Corpus Pragmatics refers to "studies of language use that employ large, computer-readable collections of language" (Jucker, 2013). The core of pragmatic investigations lies in the use and the functions of language. Form-to-function investigations examine particular linguistic entities and their functions. Conversely, function-to-form investigations focus on specific functions and the range of linguistic devices employed to perform them. Form-to-function research questions lend themselves more easily to corpus linguistics investigations while function-to-form ones are far more challenging as they do not automatically turn into search strings. Examples of form-to-function (Aijmer, 2002, 2008, 2009, Culpeper & Kytö, 2010) and function-to-form investigations (Aijmer, 1996, Jucker & Taavitsain, 2014) were provided.

Another combined approach (O'Keeffe & Walsh, 2012) to the study of spoken discourse was presented. Although CA and Corpus Linguistics have different origins and focus on different research objects, they share methodological similarities: they use empirical, naturally occurring data; they refer to baseline comparisons with other types of interactions; they investigate language in its social context; analyses are carried out at turn level for a better understanding of context. Thus, CA and Corpus Linguistics can be used as complementary approaches (O'Keeffe & Walsh, 2012). The term "CLCA" refers to this conjoint approach, which offers a more "up-close" description of spoken interactions in context than would have been normally provided if each approach had been employed on its own. There is a current trend to use smaller, specialised and highly context specific corpora that are frequently based

on spoken interactions. Such corpora are particularly suitable for a conjoint approach, which allows the investigation of other discourse features.

The empirical study of this thesis has the objective of investigating some of the key discourse, pragmatic and interactional phenomena comprising conversational competence (proposed in section 2.3.2) in the oral production of Brazilian learners of English at CEFR B1. Research questions were initially elaborated bearing in mind functions that are likely to be performed in the management of discourse, in the negotitation of illocutionary meaning and in conversational practices. Then, research sub-questions were further elaborated so as to narrow the scope of investigation to prototypical discourse, pragmatic and interational phenomena, known to perform such pragmatic functions and which lend themselves to corpus investigations.

This study included the participation of twenty Brazilian learners of English at CEFR B1+, who attended a general English course at a language institute in the South of Brazil between the second semester of 2012 and the second semenster of 2013. Before starting the data collection phase, the project of this study was submitted to CEP-PUCRS for approval. Once permission had been granted, the subjects of this study were informed of the project aims and of procedures and signed consent forms. In order to collect samples of learners' oral production, a public version of the FCE speaking test was selected, based on three main reasons. Firstly, as a formal assessment instrument, the FCE speaking test is designed and validated to give candidates the opportunity to produce enough samples of language so as to have their speaking ability assessed in four analytical criteria. Secondly, as it is delivered in a paired format, the test gives candidates the opportunity to display their conversation management ability. Thirdly, learners' spoken performance in tasks whose level of linguistic difficulty is slightly higher (CEFR B2) than that of their current level (CEFR B1 +) is likely to reveal the most obvious challenges faced by them in terms of moving beyond the intermediate plateau. Their performance may also highlight conversational phenomena they already feel confident with.

The Corpus Query Processor web server (CQPweb) at Lancaster University, UK, (https://cqpweb.lancs.ac.uk/was) was used for access to benchmark corpora and for native speaker quantitative and qualitative analyses. As the subjects of this investigation had been

exposed to textbooks and classroom materials based mostly on the British variety, it felt methodologically adequate to focus on British English native speaker corpora available on CQPweb for comparisons. The analysis options available in CQPweb include concordancing, collocations, distribution tables, frequency lists and key words. As CQPweb does not allow users to upload their own corpora, the free software "AntConc 3.4.1" (Anthony, 2014) was used for learner data analyses. Word List, Concordance and File View were the tools used in the analyses of the Brazilian learner corpus.

Three corpora constituted the core of empirical investigations. The spoken sub-corpus of the BNC Sampler contains 1,140,055 words distributed in 98 texts and was selected as one of the benchmark corpora of this study. However, as it includes samples of spoken language in different genres, a more specialised corpus of spoken conversation was compiled from another spoken corpus. The Diacronic Corpus of Present-day Spoken English contains 1,227,063 word spread in 280 texts and it portrays relatively contemporary spoken English, covering a period of a quarter of a century, from the 1960s and early 1990s. DCPSE corpus metadata information was used as a filter in the compilation of a specialised corpus named as "face-to-face conversations from the 1990s". This spoken sub-corpus contains 271,360 words distributed in 110 texts and was also used as a benchmark corpus.

As for the Brazilian data, a small corpus comprising 10 FCE interviews was built. 2 hours 40 minutes and 14 seconds of oral speech were manually transcribed following the conventions for transcribing audio files suggested by O'Keeffe, McCarthy and Carter (2007, p. 6), which included extra-linguistic information, filled pauses and backchannels. Learners' oral production was reproduced as originally uttered by them and no attempt was made to correct their grammatical, lexical, phonological and pragmatic mistakes. The final outcome was a small specialised corpus of 23,803 word tokens, displaying Brazilian learners' oral production at CEFR B1 +. Comparisons of relative frequencies derived from the three corpora were normalised to 10,000 due to the differences in corpus size.

In addition to the above three corpora, other native speaker and learner corpora were referred to in chapter sub-sections for specific purposes, namely "The Louvain International Database of Spoken English Interlanguage" (LINDSEI), "The Louvain Corpus of Native English Conversation" (LOCNEC), "The Bergen Corpus of London Teenage Language" (COLT), "The London-Lund Corpus of Spoken English" (LLC), "The Cambridge and Nottingham

Corpus of Discourse in English" (CANCODE), "The Limerick Corpus of Irish English" (LCIE) and "The Corpus of Contemporary American English" (COCA).

Chapters 4, 5 and 6 address the research sub-questions of this thesis. Chapter 4 aims at specifying the most common discourse marking adverbs used to mediate segments of discourse in conversations by Brazilian learners at CEFR B1. Chapter 5 focuses on the most common explicit and implicit adverbial hedges used to mitigate representative speech acts. Chapter 6 describes the most common minimal response tokens used to express good listenership. Each of these chapters discusses its findings in the light of specific research sub-questions. The results are intended to provide evidence of pragmatic markers that seem to pose a challenge to Brazilian learners of English at CEFR B1. Let us now turn our attention to the empirical investigations and their findings as presented in the next 3 chapters.

\

4. MEDIATING SEGMENTS OF DISCOURSE IN CONVERSATIONS: DISCOURSE MARKING ADVERBS

This is the first of three chapters that address the research sub-questions of this thesis by presenting the results and discussion based on the application of the methodology proposed in the previous chapter. All of these chapters are structured in a similar fashion. I initially introduce the chapter and detail its structure. Then I present a brief review of the literature related to the micro-elements under investigation. Next, I detail the main findings from the data in the Brazilian learner corpus and in native speaker corpora. Lastly, I compare and discuss these findings in the light of the specific research sub-question and make recommendations.

This chapter covers the following research question, focusing specifically on its sub-question:
RQ1: *How do Brazilian learners at CEFR B1 mediate segments of discourse in conversations?*
RSQ1: *What are the most common discourse marking adverbs used to mediate segments of discourse in conversations by Brazilian learners at CEFR B1? How do their frequency of use and pragmatic functions compare to those of native speakers?*

In the first section of this chapter, I briefly present corpus-based views on discourse marking adverbs (Carter & McCarthy, 2006) and stance adverbs (Conrad & Biber, 2000). I then list the most common adverbs in the Brazilian learner corpus which seem to operate as discourse markers and justify my choices for empirical investigations, namely the adverbs "*well*", "*really*" and "*actually*". Next, I compare their frequencies to those in benchmark corpora. Lastly, I detail the most common functions associated with these adverbs (Jucker, 1993, Biber et al., 1999).

In the sub-sections of this chapter, I investigate the pragmatic functions of each of the selected adverbs in native speaker and learner corpora and detail frequencies. I introduce each sub-section with a categorisation of the pragmatic functions of the item under investigation (Aijmer, 2011, Paradis, 2003, Aijmer, 2002) and related corpus-based findings. Next, I present the Brazilian learner corpora data (position, frequency and functions), qualitative analyses and comparisons with other corpora. I also include occasional concordance lines

from the benchmark corpora so as to illustrate less frequently observed functions or collocations in the Brazilian data.

In the final part of this chapter, I summarise the main points from the previous sections, discuss findings in the light of the proposed research sub-question and make some pedagogical recommendations.

4.1. DISCOURSE MARKING ADVERBS

In section 1.1.3, different views on the characterisation of discourse markers (Halliday & Hasan, 1976, Hölker 1991, Brinton, 1996, Carter & McCarthy, 2006) were detailed. It was stated that discourse markers link segments of discourse in ways which reflect choices of monitoring, organisation and management by speakers (Carter & McCarthy, 2006). It was also mentioned that among the most common and frequent markers that facilitate openings and closings are single words like "*so"*, "*(all) right"*, "*right then"*, "*now"*, "*good"*, "*well"*, "*okay (then)"* , "*anyway"* and "*fine"* (Carter & McCarthy, 2006, p. 214). According to Carter and McCarthy (2006), most discourse markers belong to the general class of adverbs. Nonetheless, it is their function as organisers of larger stretches of discourse that qualifies them as such.

With regards to the different types of meaning that adverbial phrases can convey, Carter and McCarthy (2006, p. 456) indicate the following categories: manner, place, time, duration, frequency, degree, focusing, modal, evaluative, viewpoint and linking. Whilst manner, place and time adverbs are fully integrated in the clause, other adverbs are less integrated and may modify the whole sentence or utterance. Typical examples of adverbs that are located peripherally are evaluative and viewpoint adverbs, which are also referred to as disjunct adverbs. Conversely, linking adverbs that express a logical relationship between two clauses of sentences are referred to as conjunct adverbs. Let us examine this position distinction through the examples below (Carter & McCarthy, 2006, pp. 456-458):

1. "*She blushed* ***furiously"*** (adverb modifying the manner in which she blushes; integrated in the clause)
2. ***"Personally****, I think this is a waste of time".* (disjunct adverb expressing a viewpoint; peripheral position)

3. *"The bad news is that your shares are almost certainly overvalued at present, as the entire bank sector. The good news is that they are probably worth keeping* ***anyway****".* (conjunct adverb expressing a logical relation of contrast; linking two clauses)

In the above sentences, "*personally*" in example 2 functions as a discourse marking adverb since it is displayed in a peripheral position, not adding any propositional content to the clause but expressing speaker attitude towards what is being said. By contrast, the adverb "*anyway*" in example 3 conveys a dismissive adversative meaning (*no matter under which, or what, circumstances)* falling into the categorisation of formal markers of cohesion.

However, as previously mentioned in section 1.1.3, "*anyway*" may also have a cohesive meaning (*let's get on with the job*). Therefore, the adverb "*anyway*" is used as a discourse marker in the example below (Halliday & Hasan, 1976, p. 270):
"They changed over to a most peculiar kind of train which you don't see now. I've forgotten what it is called. Was it called a "steam coach"? ***Anyway*** *it was just one coach but it ran by steam and it made a funny noise".*

As for stance adverbs in particular, Conrad and Biber (2000) report on the findings of their corpus-based investigation on the different ways in which speakers and writers use adverbials to mark their personal stance. Conrad and Biber (2000, p. 57) refer to "stance" as a cover term for the expression of personal feelings and assessments in three major domains:
1. Epistemic stance[42]: commenting on the certainty (or doubt), reliability, or limitations of a proposition, including comments on the source of information;
2. Attitudinal stance: conveying the speakers' attitudes, feelings, or value judgements;
3. Style stance: describing the manner in which the information is being presented.

Epistemic stance adverbials are employed to perform the following functions (Conrad & Biber, 2000, pp. 59-60):

a) Indicating the degree of certainty or doubt concerning the proposition, e.g., "*Well,* ***perhaps*** *he is a little bit weird*".
b) Commenting on the reality or actuality of the proposition: "*You can* ***actually*** *hear what she's saying*"; "*You're wise to lock it* ***really***";

[42] Epistemic stance is addressed in chapter 5 within mitigation.

c) Indicating that the proposition is somehow imprecise: "*It seems to clean it up **if you call it that***";
d) Indicating the source of information, either specifically or by implication with words such as "*apparently*" and "*evidently*": "*They just **apparently** built up huge quantities of dry bird droppings and these were staggering high*".
e) Marking limitations of the information or identifying the perspective from which the proposition is *true: "**From our perspective**, movement success is paradoxical".*

Attitudinal stance adverbials encompass a wide range of meanings, conveying attitudes, feelings, value judgements, or expectations. Conrad and Biber (2000, p. 60) present some examples of attitudinal adverbs:
"*…but **fortunately** I put it in a folder so the folder was destroyed*".
"***Amazingly** Adam walked away from the crash with just a graze on his left shoulder*".

Style stance adverbials comment on the manner of speaking, in other words, they comment on the way in which information is being presented or meant to be understood. Examples of style stance adverbials include (Conrad and Biber, 2000, p. 60):
*"**Honestly**, I've got a headache".*
*"If his- desires were carried out, we'd, we'd be talking about thousands of pounds. **Literally**".*

Furthermore, Conrad and Biber (2000) indicate that the stance adverbials "*probably*", "*actually*", "*really*" and "*sort of*" are particularly common in conversation, each occurring more than 60 times per 100,000 words. If taken together, they account for 70 per cent of all epistemic stance markers of conversation. In addition to marking the speaker's stance, these adverbials have important social functions. While markers of doubt such as "*perhaps*" and "*probably*" are used for making suggestions, markers of actuality such as "*actually*" and "*really"* are used for softening disagreements.

Now that we have discussed the functions of adverbs, discourse marking adverbs and stance adverbs, let us investigate the adverbs produced by the subjects of this study. I used AntConc's word list tool to view the rank positions and raw frequencies of all the words in the Brazilian learner corpus. As the corpus is not annotated, I made a manual list of all the

adverbs produced by the Brazilian learners and indicated their raw frequencies (actual number of occurrences) and relative frequencies (per 10,000 words) as shown in Table 1. The rank position was not included since the lexical items had been ranked alphabetically according to the number of occurrences.

Table 1: List of adverbs in the Brazilian learner corpus

Adverb	Raw frequency	Relative frequency
maybe	**72**	**30.2482**
really	**47**	**19.7454**
recently	23	9.6626
just	**20**	**8.4023**
well	**14**	**5.8816**
probably	**13**	**5.4614**
always	9	3.7810
especially	9	3.7810
usually	8	3.3609
hard	6	2.5206
nowadays	6	2.5206
actually	**4**	**1.6804**
carefully	3	1.2603
exactly	3	1.2603
friendly	3	1.2603
normally	3	1.2603
already	2	0.8402
possibly	2	0.8402
totally	2	0.8402
absolutely	1	0.4201
early	1	0.4201
loudly	1	0.4201
principally	1	0.4201
quickly	1	0.4201

As highlighted in Table 1, the adverbs "*maybe*", "*really*", "*just*", "*well*", "*probably*" and "*actually*" seem to be the most common discourse marking adverbs used by the subjects of this study. "*Recently*" was not considered since almost all of the occurrences had been produced by the interlocutor when delivering the rubrics of the speaking test. Manner adverbs, frequency adverbs and adverbs with fewer occurrences than 4 were also disregarded. For the purposes of this sudy, I chose to investigate the pragmatic functions of "*really*", *"well*" and

"*actually*" in this chapter since these adverbs are commonly used to mediate segments of discourse with regards to the different parts of discourse and also to the relations between speakers and parts of discourse. While "*well*" is a common and frequent single word adverb used to facilitate opening and closings of conversations (Carter & McCarthy, 2006), "*really*" and "*actually*" are highly frequent stance adverbials used to mark speakers' attitude or stance to the message (Conrad & Biber, 1999). The adverbs "*maybe*", "*just*" and "*probably*" are traditionally associated with mitigation and will be addressed in chapter 5.

Let us now investigate the frequencies of "*really*", "*well*" and "*actually*" in the native speaker benchmark corpora. I used CQP's frequency list tool to display the frequency of word forms. Table 2 summarises the rank position, raw frequencies and relative frequency (normalised to 10,000 words) of "*well*", "*really*" and "*actually*, functioning both as regular adverbs and as discourse markers, in the two benchmark corpora of this study, namely the "BNC Sampler" spoken sub-corpus (1,140,055 words in 98 texts) and the specialised sub-corpus "face-to-face conversations from the 1990's" (271,360 words in 110 texts).

Table 2: "*Well*", "*really*" and "*actually*" ranking and frequencies in the benchmark corpora

Well	Rank position	Raw frequencies	Relative frequency
BNC spoken	32	6,200	54.3833
Face-to -face conversations	27	1,713	63.1264

Really	Rank position	Raw frequencies	Relative frequency
BNC spoken	100	1,750	15.3501
Face-to -face conversations	65	788	29.0389

Actually	Rank position	Raw frequencies	Relative frequency
BNC spoken	129	1,177	10.324
Face-to -face conversations	86	515	18.9784

The above data attest high relative frequencies of "*well*", "*really*" and "*actually*" in both corpora, indicating ever higher rank positions and relative frequencies in the specialised sub-corpus of face-to-face conversations, and thus, strengthening their association with face-to-face conversations. As "*well*" is also commonly used as a manner adverb (i.e. "*She can dance Salsa very **well**"),* a further investigation was carried out. After conducting thinned queries of "*well*" in both corpora, the manual analyses indicated few occurrences of "*well*" as a manner

adverb. Out of 100 occurrences in each corpora, the BNC Sampler spoken sub-corpus displayed only 6 occurrences of "*well*" as a manner adverb whereas 5 occurrences were found in the face-to-face conversations from the 1990s sub-corpus. These results corroborate the assumption that "*well*" is much more commonly used as a discourse marker. Another point to add is that "*well*" was more frequent than "*really*" in British English native speaker production while the Brazilian learners of this study produced "*really*" more frequently than "*well*".

Now let us present what the literature says about the most common functions of "*well*", "*really*" and "*actually*". Jucker (1993) accounts for the pragmatic uses of "*well*" within the Relevance theoretical [43] framework. For Jucker (1993), the discourse marking adverb "*well*" has four main uses, which can be related to one core meaning:

1. It can be used as a marker of insufficiency, indicating some problems on the content level of the current or proceeding utterance;
2. It can be used as a face-threat mitigator, indicating some problems on the interpersonal level;
3. It can be used as a frame marking device indicating a topic change or introducing direct reported speech;
4. It can be used as a delay device.

In all the above uses, "*well*" is "some kind of signpost, directing the way in which the following utterance should be processed by the addressee" (Jucker, 1993, p. 438). Utterances are processed against a background, i.e., "against the background of the addressee's personal cognitive environment and against the background as created by the previous discourse" (Jucker, 1993, p. 438). Thus, the discourse marking adverb "*well*" signals that the addressee needs to reconstruct the background against which the upcoming utterance can be processed.

From a corpus-based perspective, Biber et al. (1999) point out that "*well*" is a versatile discourse marker though it "appears to have the general function of a deliberation signal, indicating the speaker's need to give (brief) thought or consideration to the point at issue". It is a common turn indicator serving to relate a speaker's response to the ongoing conversation.

[43] Sperber and Wilson's Relevance Theory (1986, 1995) was discussed in Corsetti (2009, pp. 38-45)

Let us examine the following examples of the different functions of "*well*" (Biber et al., 1999, p. 1087):

1. **Responsive role:**

   "*He said, **well**, I'd like to read a little bit about it first – I said, **well**, can you read my first volume, you will see for yourself, you know, but he said, **well**, what about this, **well**, you know, **well**, you know, and he hems and haws and he won't come out and say yes or no*".

2. **Response marker following an agreement marker (e.g. "*yeah*"):**

   a) "We were talking about last night".

   *b) "**Yeah**, **well**, I used to walk a lot, but I, er, I <…> now all I do is eat".*

3. **Marking continuation but with something of a contrast:**

   a) "You are always hungry".

   *b) "**Well**, I'm not now".*

4. **Signalling self-correction or deliberation over a choice of expression (occurring in the middle of an utterance:**

   *"The boss and the secretary work late all night, **well**, not all night but late into the night*".

With regards to "*really*", Biber et al. (1999) state that both British and American conversations display "*really*" as a modifier of a variety of adjectives, particularly with the four most common collocates "*good*", "*nice*", "*bad*" and "*funny*" (all occurring over 20 times per million words) . "*Really*" can be also used as an amplifier in conversations (Biber et al., 1999, p. 564), for instance: "*He'll look **really** [sweet]*". Nevertheless, "*really*" is also a stance adverbial. Stance adverbials perform the function of commenting on the content or style of a clause or a particular part of the clause. Biber et al. (1999) indicate that certain stance adverbials can have ambiguous or multiple functions and that "*really*" is a particularly difficult adverb to analyse. Certain examples of "*really*" have an epistemic stance meaning "in reality" or "in truth", in particular in initial and final positions (Biber et al., 1999, p. 857): *"**Really** you've noticed the difference?"* and *"I had no choice **really**".* When it occurs medially, it can also have an epistemic stance meaning with propositions that concern absolute characteristics (Biber et al., 1999, p. 858): *"Was Molly ever **really** alive alive-oh?"* and *"<…> it must be admitted that we have no proof that plants **really** do obtain all their carbon dioxide in this way".* However, in medial position with gradable propositions, the

determination of the meaning is ambiguous, for instance: "*It's really wonderful*" – either a stance meaning of "in reality" or as an adjective intensifier.

As for "*actually*", Biber et al. (1999, p. 870) indicate that conversations, if compared to other registers, display high frequencies of adverbials marking actuality such as "*actually*" and "*really*". Speakers make a point of identifying propositions as factual or real, for instance: "*I've **actually** got very strong teeth*". In addition, Carter and McCarthy (2006, pp. 28-29) state that "*actually*" can be used to perform the following functions:

1. **Marking that something is contrary to expectations, used emphatically:** "*He **actually** admitted that he enjoyed it*".
2. **Implying a contract between a desirable and an undesirable situation:** "*So, here is a practical seminar that **actually** offers solutions to the challenges women managers face".*
3. **Signalling topic openings, contrasts in topics, specifying within topics, used as a discourse marker**: *"Could you tell me where your manuals are kept? **Actually** I'm looking for a Haynes manual".*
4. **Hedging statements, making them less direct or threatening:** "*We had an argument **actually**, a few months ago".*

To sum up, in this section we have covered discourse markers and stance adverbs and presented data retrieved from the Brazilian learner corpus and the benchmark corpora. While most discourse markers belong to the general class of adverbs, it is their function as organisers of larger stretches of discourse that qualifies them as such (Carter & McCarthy, 2006). Stance adverbs mark speakers and writers' stance, i.e. the expression of personal feelings and assessments in three major domains: epistemic stance, attitudinal stance and style stance (Conrad & Biber, 2000). The adverbs "*maybe*", "*really*", "*just*", "*well*", "*probably*" and "*actually*" seem to be the most common discourse marking adverbs used by the subjects of this study. The adverbs "*really*", *"well*" and "*actually*" were chosen for investigation in this chapter since they are commonly used to mediate segments of discourse with regards to the different parts of discourse and also to the relations between speakers and parts of discourse. Native speaker data attest high relative frequencies of "*well*", "*really*" and "*actually*" in both benchmark corpora, indicating ever higher rank positions and relative frequencies in the specialised sub-corpus of face-to-face conversations, and thus,

strengthening their association with face-to-face conversations. The adverbs "*maybe*", "*just*" and "*probably*" are traditionally associated with mitigation and will be addressed in chapter 5.

Let us now investigate the pragmatic functions "*well*", "*really*" and "*actually*" in specifc contexts.

4.1.1. "*Well*"

This section initially details Aijmer´s (2011) categorisation of the functions of "*well*" and her corpus-based findings vis-à-vis learner (Swedish component of LINDSEI) and native speaker (LOCNEC) spoken production. It then presents Brazilian learner corpora data (position, frequency and functions), qualitative analyses and comparisons with other corpora. It also includes concordance lines (BNC Sampler spoken-sub-corpus) so as to illustrate less frequently observed functions or collocations in the Brazilian data.

Aijmer (2011) compares the uses of "*well*" in the Swedish component of the learner corpus LINDSEI (Louvain International Database of Spoken English Interlanguage) and in its native speaker counterpart LOCNEC[44]. Aijmer (2011) suggests that advanced EFL learners use frequent pragmatic markers like "*well*" though such uses differ from native speakers' ones. Through the comparison between the distribution of "*well*" in both corpora, Aijmer argues that the Swedish learners overused "*well*" in comparison to the native speaker figures: in the Swedish component of LINDSEI, "*well*" occurred 544 times per 100,000 words and, in LOCNEC, 421 times. She also indicates that "*well*" was used in initial, medial and final positions by both learners and native speakers but with different frequencies.

In relation to the different positions, Aijmer (2011) states that "*well*" initial is usually associated with turn-taking and the organisation of discourse. Native speakers used "*well*" to signal turn-taking more than learners did. By contrast, learners produced "*well*" more in medial position. This overuse of "*well*" in medial position is consistent with the hypothesis that learners use "*well*" all over the place, with no special strategy in mind, rather than for interpersonal functions (Aijmer, 2011, p. 234). Learners sometimes placed "*well*" in final position, reflecting the fact that they did not know how to continue the utterance.

[44] LOCNEC (Louvain Corpus of Native English Conversation) replicates LINDSEI tasks but has young native speakers of British English as interviewees instead of learners. It contains 170,553 words (Aijmer, 2011).

Aijmer (2011) analysed "*well*" as primarily a mental state interjection with discourse function, which can be associated with the speaker's deliberation. As previously mentioned, Biber et al. (1999, p. 1086) state that "*well*" "appears to have the general function of a deliberation signal, indicating the speaker's need to give (brief) thought or consideration to the point at issue". However, Aijmer (2011) argues that the presence of "*well*" may have little to do with consulting one's thoughts but be associated with hesitation or pausing and the speaker's difficulties in the communication situation (speech management problems). Aijmer (2011) adopts the following typology of "*well*" in her comparisons between native and learner corpora:

1. Speech management: choice, change, prospective (introducing a new turn), marking stages in a narrative and quotative[45];
2. Attitudinal: opinion and disagreement.

Concerning speech management, "*well*" co-occurring with a mid-utterance break can be associated with two functions: the choice-related (used for pausing and planning) and the change-related (or repair) function. The choice-related function enables speakers to gain time and plan ahead (Aijmer, 2011, p. 237):

(1)

<B> <overlap/> and they mm erm. Took me out to parties and <breathes> **well** different places so I could meet more people…so em… to get over my shyness <B/>

(LINDSEI-SW 014)

It may also be preceded and followed by (other) filled or unfilled pauses as a turn-holding device while the speaker is searching for words (Aijmer, 2011, p. 237):

(2)

<B> and he seems to be: .**well**. viewing her and he's drawn some: sketches some outlines on the on the canvas and ha= hasn't really done very much work <B>

(LINDSEI-SW 005)

"*Well*" co-occurred more frequently with pauses in the learner corpus while it occurred more frequently without a pause in the native speaker corpus. Both learners and native speakers used combinations of filled and unfilled pauses and other markers with "*well*" but their

[45] The functions of "marking stages in a narrative" and "quotative" will not be detailed as the subjects of this study did not engage in speaking activities that required these functions.

choices differed: learners used the combinations "*erm*", "*er*", "*eh*", "*em*" more often while native speakers used "*yeah well*" more frequently.

As for the change-related function, "*well*" can co-occur with self-interruption and a re-start; it is also used as a clarification or correction (in particular together with "*I mean*" as a way for the speaker to make a clarification before continuing (Aijmer, 2011, p. 240):

(3)

and: even though h= the painter thought it was a good picture and that it was a picture that really that suited her g= **well I mean** he thought that was how she really looked so now he:had to: he had to make it more 1= her to look more beautiful

(LINDSEI-SW 010)

In the prospective function, when "*well*" occurs in initial position, it is used as a feedback marker (with the meaning deliberation or pausing) followed by a new turn. "*Well*" was also employed by both learners and native speakers to perform a task provided by one of the set topics (Aijmer, 2011:242):

(4)

<A> okay well tell me about it <A/>

<B> **well** I thought. I suppose it's a little bit personal but I thought I might tell you about something that happened in the first year <B/>.

(LOCNEC E07)

Aijmer (2011) indicates that the boundary may be marked by "*well*" in combination with other markers such as "*okay*", "*now*", "*right*", "*all right*". In addition, "*well*" in an answering move or introducing one of the set topics occurred 55 times in LINDSEI-SW and almost twice as often in LOCNEC with 107 instances.

In the second broader categorisation, the attitudinal "*well*" expresses attitudes or feelings to the hearer or to the preceding discourse. In the below example, "*well*" has an attitudinal function associated with downtoning and politeness, mitigating the speaker's opinion expressed by "I'm not sure I think" (Aijmer, 2011, p. 246):

(5)

<A> erm..is that generally positive. nowadays </A>

<B> **<breathes> <sighs> well I'm not sure I think** it's it's..it's both I can't really decide what people around me think </B>
(LINDSEI-SW 006)

"*Well*" in the collocations "*well no*", "*well actually*", "*yeah well*", "*well but*" and "*well yes but*" indicates a correction, explanation or clarification if the preceding discourse is challenging or makes a wrong assumption. "*Well*" can also introduce the answer to a question that requires confirmation, clarification, explanation or justification. More typically, it is associated with the speaker's unwillingness or reluctance to answer a question directly. "*Well*" may also be used when an opinion is qualified in some way, e.g. "*Well, I think*".

Overall, Aijmer's results (2011, p. 248) demonstrate that the Swedish learners used "*well*" significantly more for speech management than the native speakers did and less for attitudinal functions. Learners overused "*well*" in the choice-related function but not in the change-related function. Native speakers, on the other hand, used "*well*" significantly more in the change-related and prospective functions. As for the gap between learner and native speakers in the attitudinal functions, Aijmer (2001) concludes that it is important to discuss the uses of "*well*" as a means to express feelings and attitudes in the classroom. Learners may not be aware of its interpersonal function and its importance in establishing and maintaining good relationships in the communication.

Let us now focus on the Brazilian learner data. Overall, there were 14 occurrences of "*well*" distributed in 4 interviews. However, 5 occurrences out of 14 were preceded by "*as*" forming the adverbial phrase "*as well*". Table 3 summarises the positions and the functions of the 9 occurrences of "*well*" as a one-word pragmatic marker. The percentage column indicates the distribution among functions.

Table 3: Position and functions of "*well*" in the Brazilian learner corpus

Well	Utterance initial	After filled and unfilled pauses	Medial position	Final position	Percentage
Choice related function 4 occurrences		2	2		44.5%
Change related function 2 occurrences	1	1			22.2%
Opinion function 3 occurrences	2	1			33.3%
9 occurrences total	3	4	2		100.0%

Considering the above data, the Brazilian learners of this study employed "*well*" mostly for speech management functions, namely choice and change-related functions. Attitudinal functions were less frequent with 3 occurrences for expressing opinions. These data are aligned with Aijmer's learner results (2011), corroborating the assumption that learners use "*well*" more frequently for speech management functions than native speakers do (LOCNEC data) and less for attitudinal ones. The subjects of this study also used "*well*" more frequently in the choice-related function than in the change-related function. Instances of "*well*" occurred in three contexts. Preceded by filled and unfilled pauses in utterance initial was the most common position, followed by utterance initial and medial positions, respectively. "*Well*" preceded by filled and unfilled pauses strengthens Aijmer's assumption (2011) that learners may use it more as a hesitation device rather than as a marker of deliberation. Let us now illustrate the above results with corpus-based data.

In the choice-related function in initial position, learners used "*well*" preceded by filled or unfilled pauses as a turn-holding device while searching for words or planning ahead as in the examples below:

(6)

<S1> Okay, thank you. Now … here are your photographs. They show people of different ages on educational visits. I'd like you to compare the photographs and say what you think the people will learn on their visits.

<S11> **Okay, uh well** we have a pictures of old people in a museum and I think this interesting because we always uh are, we, I think we need to be all the times open to learn…

(FCE P1- B1 BR-05)

(7)

<S10> … But if the place is good you, you can pay for that if you're <SE> pause</SE>, but not so expensive.

<S11> Yes, I agree**. And <SE> long pause</SE> well**, so we…

(FCE P3- B1 BR-05)

In example (6), "*well*" is preceded by the pragmatic particles "*okay*" and "*uh*" and it is used a means for gaining time and planning ahead. In example (7), "*well*" is preceded by a long unfilled pause and it is also used as a turn-holding device. These occurrences reinforce the above hypothesis that learners tend to use "*well*" mostly as a hesitation device.

(8)

<S1> What do you like about living in Porto Alegre …?

<S19> I like the….

<S1> What about you?

<S18> **Uh, well I, I**, I like very much Porto Alegre, uh I live here since uh, uh nineteen ninety seven and <SE> pause /SE> and I, I gets uh useful with the city and uh I learned to, to, to love uh, uh the city and I, I like very much the, the big events, cultural events, I liked the museums, uh the many <SE> pause /SE> uh cinemas rooms <SE> laughing /SE> **uh and <SE> pause /SE>, well, uh** part of my family live here and I have a lot of friends here and now I'm, I'm, I'm far away, I'm a little far away from Pelotas uh nowadays.

(FCE P1- B1 BR-09)

The extract above (8) displays two examples of "*well*" used for different purposes. The first occurrence of "*well*" is preceded by a filled pause "*uh*" and used to initiate a long turn in which speaker 18 is expected to justify why he likes living in Porto Alegre. Thus, it is used to introduce the speaker's opinion. However, "*well*" is followed by two false starts, conveying that subject 18 has mixed feelings about Porto Alegre. The second occurrence of "*well*" is preceded by a number of filled ("*uh*") and unfilled pauses produced within the previous long turn. "*Well*" is in medial position performing the choice-related function. It is used for

pausing and planning what to say next. The continuation of the utterance shows an abrupt change to the information structure.

In the change-related function, learners used "*well*" in initial position as a clarification or correction. In the example below (9), subject 11 produces a long turn in which she compares the different reasons for old and young people to enjoy going to the seaside. The ideas she conveys make the hearer laugh. Thus, in order to mitigate the seriousness of her statement she uses "*well*" as a means to clarify what she means, e.g. she is referring to young and old people in Brazil. It is interesting to notice the co-occurrence of the Brazilian particle "*né*", which has the same function of a question tag, i.e. seeking confirmation:

(9)
<S11> The places that they go? I think young people likes to go to the beach. And the, but the beach, beach, really close to the ocean, you know, and the waves. And the old people like to go to the beach also, but not to stay in the beach. Nowadays they like to stay in the house, in the morning go to walk in the, in the beach and return to the house and stay there <SE> laughing</SE>.
<S10> <SE> chuckles</SE>
<S11> And the, or they like to go to a, to a countryside, I think the old people, but the youngs, the youngest likes to go to the beach. **Well, here, né**?
(FCE P4- B1 BR-05)

Another occurrence of attitudinal "*well*" was produced in initial position, without any hesitation and followed by the opinion marker "*Well, I think*", thus, conveying assertiveness and qualifying one's opinion:
(10)
<S1> …First talk to each other about how successful these suggestions might be. Then decide which two would attract most people.
<S11> **Well, I think** the TV will not attract.
(FCE P2- B1 BR-05)

Overall, a salient feature highlighted through the Brazilian data analysis was the uneven distribution of "*well*" as a one-word pragmatic marker among interviews and subjects. Out of the 10 interviews, "*well*" occurred in 3 texts only. In addition, out of the 20 subjects, only 3

participants included "*well*" in their oral production (subject 11: 5 occurrences; subject 18: 3 occurrences, subject 21: 1 occurrence. This possibly indicates lack of familiarity with the discourse functions of "*well*" by the Brazilian learners of this study, and thus, it may signal underuse.

Furthermore, "*well*" was mostly used within speech management functions and often preceded by filled and unfilled pauses in initial position (4 occurrences) or in medial position (2 occurrences). These results corroborate the assumption that "*well*" was mostly used as a floor holding device, similarly to Aijmer's results (2011). In parallel, all interviews displayed a far higher frequency of "*uh*" as a hesitation device: 836 occurrences. "*Uh*" was employed by subjects mostly in initial and medial positions as a means to gain time and plan ahead and also when searching for language. The interviewer also used "*uh*" in initial position while deciding on what questions to ask to individual subjects. It is possible to infer that while some subjects of this study used "*well*" as a turn-holding device, all subjects of this study resorted to "*uh*" as a hesitation device, which frequently occurs in Brazilian Portuguese too, possibly indicating language transfer.

Aijmer's (2011) results already account for native speaker data. Thus, frequencies from the reference benchmark corpora of this study were not included in this section. However, as the Brazilian learner results indicate some possible underuse of "*well*" as a discourse marker, concordance lines retrieved from the BNC Sampler sub-corpus have been included for illustrative purposes. Figure 4 presents collocations of "*well*" followed by the pronoun "*I*" performing mostly attitudinal functions.

hear any more ? Mary lives near them . Well then ?	Well I	'm rather curious erm phone call from Eric , Ernest Ernest is
the travel is going to be something is n't it ? Erm	Well I	think who 's , who 's thinking of going ? I ,
Islington . Monday mm . Would you be interested in it ?	Well I	might this , this , the time fitting in with the families
be a good idea to have She 's taken someone 's .	Well I	think , we 'll , we 'll see what , we 'll
same thing , yeah . Are you going to , interest .	Well I	called a meeting but I did n't think I advertised it well
n't , I do n't think we can always do it .	Well I	'm , yeah , ern I wondered about sort of Executive Committee
nearer then sit here yes , ok , is that alright ?	Well I	think I 'll start off by telling you a little bit about
people it might be , yes . I I Bloody hell !	Well I	think , you know That 's that 's not very good English
no respect for your ! Could we have punctuality down there ?	Well I	was goi I would quite T like that . Right . Colour
Mm mm . I 'm just asking Testing us . why ?	Well I	have n't , I have n't got an axe to grind !

Figure 4: Extract of concordance lines for "*Well*" + "*I*" in the BNC Sampler spoken sub-corpus

As shown in the concordance lines above, "*Well I'm rather curious*" expresses the speaker's attitude. Examples of "*Well I think*" seem to indicate the use of "*well*" as a constituent of a fixed expression to introduce one's opinion. "*Well I might*" indicates some uncertainity. "*Well I was goi…*" and "*Well I haven't*" introduce justifications. "*Well*" followed by two filled pauses "*yeah*" and "*erh*" as in "*Well I'm, yeah, erh …*" conveys relunctance for agreement.

Furthermore, all the instances of "*as well*" in the Brazilian data seemed to be performing cohesive functions. Thus, some examples are reproduced in this section for illustrative purposes too. All occurrences of the adverbial phrase were produced in the same interview (by both speakers). Three occurrences of "*as well*" were placed in final position while students were performing the collaborative task. There were two occurrences of "*as well*" in medial position. Examples of "*as well*" in final position:

(11)

<S4> I think is a good idea. You can have a live music.

< S5> The free coffee **as well** <SE> laughing </SE>.

<S4> <SE> overlapping</SE> Yeah <SE> laughing </SE> .

(FCE P3- B1 BR-02)

(12)

<S5> Yeah and uh the Japanese restaurants uh people seems to like, they're always full of people <SE> pause</SE>**as well**.

(FCE P4- B1 BR-02)

In example (11), "*as well*" appears as an elliptical form of the proposition "I also think that free coffee is a good idea". In example (12), learners were previously discussing the most popular restaurants in Brazil. "*As well*" seems to reinforce the popularity of Japanese restaurants in Brazil: people like Japanese restaurants; Japanese restaurants are always full of people because people like them.

In the example below, "*as well*" is placed in medial position, used as a discourse marker and meaning "similarly to subject X":

(13)

<S1> And you…?

<S5> I'm from, from Brazil **as well** and I was born here in Porto Alegre.

(FCE P1- B1 BR-02)

Concluding, in this section we have investigated the discourse functions of "*well*" and presented learner and native speaker data. The discourse marking adverb "*well*" performs speech management functions and attitudinal functions (Aijmer, 2011). The Brazilian learners of this study used "*well*" mostly for speech management functions, namely choice-related and change-related functions. Attitudinal functions were less frequent. The Brazilian data are aligned with Aijmer's learner results (2011), corroborating the assumption that learners use "*well*" more frequently for speech management functions than native speakers do (LOCNEC data) and less for attitudinal ones. The subjects of this study also used "*well*" more frequently in the choice-related function than in the change-related function.A salient feature highlighted through the Brazilian data analysis was the uneven distribution of "*well*" among interviews and subjects. This possibly indicates lack of familiarity with the discourse functions of "*well*" by the Brazilian learners of this study, and thus, it may signal underuse. Furthermore, "*well*" was mostly used within speech management functions and often preceded by filled and unfilled pauses, corroborating the assumption that "*well*" was mostly used as a floor holding device, similarly to Aijmer's results (2011).

4.1.2. "*Really*"

This section initially details Paradis`s (2003) categorisation of the pragmatic functions of "*really*" and her corpus-based findings vis-à-vis native speaker spoken production (COLT). It then presents Brazilian learner corpora data (position, frequency and functions), qualitative analyses and comparisons with native speaker and Spanish learner data (LINDSEI). It also includes concordance lines (face-to-face conversations from the 1990s sub-corpus) so as to illustrate less frequently observed functions or collocations in the Brazilian data.

Paradis (2003) investigates the uses of "*really*" in two British English corpora. Initially, Paradis (2003) carries out semantic analyses of "*really*" in the Bergen Corpus of London Teenage Language (COLT), which forms part of the BNC and consists of half a million words of spontaneous , informal face-to-face conversation among teenagers in London. She is able to identify three functions of "*really*": truth attesting, emphasising and degree reinforcing. Following this, Paradis (2003) then conducts a further investigation of these

functions in the London-Lund Corpus as it is prosodically annotated. According to Paradis (2003), intonation also needs to be considered when interpreting "*really*". Due to the scope of this thesis, only the data from the COLT corpus will be detailed.

Paradis (2003, p. 197) investigates 1, 348 occurrences of "*really"* in COLT (26.96 relative frequency per 10,000 words) and summarises the distribution among functions as follows:

a) Truth attesting: 316 occurrences - 23%
b) Emphasising: 437 occurrences – 33%
c) Degree reinforcing: 595 – 44%

The above data indicate a higher percentage of "*really*" as a degree modifier, and thus, being used as a regular adverb. "*Really*" as an emphasiser is the second most common function, followed by "*really*" truth attester. In all three functions, "*really*" conveys pragmatic meaning.

The truth attesting "*really*" has the function of ensuring the truth of the assertion that it takes in its scope and provides implicit evidence based in "reality". It typically occurs in a separate tone unit and may be placed in all adverbial positions (Paradis, 2003, p. 198):

(1)
*"**Really** that's quite good".*
(2)
*"She loves me **really**".*
(3)
*"Sue and Bill **really** bought the farmhouse they had been dreaming of".*

In (1), "*really*" is in initial position and can be re-phrased as "in accordance with evidence from reality, that is quite good". In (2), "*really*" is in utterance final position and means "in accordance with evidence from reality, she loves me". In (3), "*really*" is in medial position and can be paraphrased as "in accordance with evidence from reality, Sue and Bill bought a house they had been dreaming of". In all the above examples, "*really*" takes scope over the whole assertion and its function is to make the listener interpret the assertion in the light of reality. Paradis (2003, p. 198) adds that in some occurrences, truth attesting "*really*" also creates a contrastive reading due to the fact that "assertive propositions come with the assumption of truth, and the explicit marking of truth by *really* tends to create a context where

the opposite is presupposed". In (3), one can infer that the listener did not expect Sue and Bill to buy the farmhouse they had been dreaming of.

In addition, "*really*" used in questions is normally a truth attester. Speakers make use of "*really*" to ask previous speakers whether what was said before is in line with reality and truth, for instance (Paradis, 2003, p. 198):

(4)

A: *"She is fucked up mentally".*

B: *"**Really?**"*

A: *"Mm I think she is mm I think she is what happened was yeah…"*

As for emphasising "*really*", it has the effect of adding subjective emphasis of situations. It is positionally constrained in the sense that it has to be placed "adjacent to a verb denoting a situation type that is attitudinal in character or alternatively a situation type that may be capable of undergoing subjective modulation in the context of a trigger element such as *really*" (Paradis, 2003, p. 199). It is often placed before the first verb and is semantically bound to attitudinal verb meanings. It is intonationally non-salient. Paradis (2003, p. 200) compares the following examples:

(5)

*"No, that's sad, that **really** is definitely".*

(6)

*"I meant to be going to the choir tonight but I **really** can't be bothered".*

(7)

*"I do actually **really** like singing".*

In the above examples, the emphasising "*really*" takes scope over situation types that are states: "that is sad"; "I can't be bothered"; "I like singing". Paradis (2003, p. 200) adds that the meaning of "*really*" is "semantically bleached as compared to the foreground and distinct REALITY notion in truth attesting *really*". Its main aim as an emphasiser is to convey speaker meaning. In example (7), "*actually*" is a factual modifier of evidence while "*really*" is a marker of epistemic subjectivity.

Furthermore, the closeness between the situation type and "*really*" is observed in its preferred combinatorial links to certain stative predicates. The most common collocating verbs with "*really*" as an emphasiser are as follows (Paradis, 2003, p. 200):

a) Don't: "*I **really** don't mind/know/care/want to/think…"*

b) Like: *"I **really** like her"*

c) Be: *"I mean he **really** is a cool guy"*

d) Want: "*I **really** want my mum to hear that*"

e) Hate: "*I **really** hate her"*

f) Hurt: "*It **really** hurts"*

Paradis (2003) indicates that all the meanings of the above predicates are attitudinal. "Really latches on to the attitudinal facet, which it emphasises, and the invited inference is that the truth of the proposition is thereby attested" (Paradis, 2003, p. 201). In other words, emphasising "*really*" is mainly schematic rather than lexical. Thus, the lexical weakening is replaced by subjective strengthening.

Paradis (2003, p. 201) clarifies that emphasising "*really*" also combines with adjectival predications (states). Such adjectives are either non-scalar as in "*I always get **really** paranoid with people I get off with*" or represent an extreme point of a scale as in "*It's **really** appalling*". Paradis (2003) also stresses that subjective emphasis only occurs in statements. In questions, "*really*" is a truth attester and takes propositional scope. However, "*really*" may be used as an emphasiser in questions where speakers express their own judgements and seek support for their opinion, for instance "*Isn't it **really** appalling*". When "*really*" is preceded by a negative element, it is a de-emphasiser, attenuating or approximating the truth of the application of the situation talked about, for instance: "*I can't **really** help it*" (Paradis, 2003, p. 202). This results in a hedged statement.

Lastly, degree reinforcing "*really*" reinforces scalar properties denoted by adjectives. Similarly to emphasising "*really*", its interpretation is grounded on mappings between concepts within the proposition. Degree reinforcing "*really*" takes scope over scalar property concepts denoted by adjectives and thus, it is also intonationally non-salient. Let us examine some examples provided by Paradis (2003, p. 202):

(8) "*Hugh is apparently **really** rude about everyone especially when he gets drunk*".

(9) "*He thinks he's* ***really*** *cool*".

In the above examples, "*really*" has the effect of reinforcing the degree of rudeness and personality, respectively. The adjectives "*rude*" and "*cool*" are based on a scale schema that enables "*really*" to develop a degree reinforcing function.

The most common scalar collocates in COLT are as follows (Paradis, 2003, p. 203):

a) Good: "*no Zed's a* ***really*** *good bloke when he's sober"*

b) Nice: "*well my dad reckons he's* ***really*** *nice anyway"*

c) Funny: "*I think she is like* ***really*** *funny like"*

d) Bad: "*he had* ***really*** *bad dandruff just now as well"*

e) Sad: "*Gran I thought your letter was* ***really*** *sad"*

f) Cool: "*I've seen Demolition man cos I'm* ***really*** *cool"*

g) Nasty: "*oh turn that off it's a* ***really*** *nasty noise"*

h) Weird: "*she's got a* ***really*** *weird accent hasn't she"*

According to Paradis (2003, p. 214), "it is argued that "*really*" is pragmatically conditioned by the speaker's wish to qualify an expression epistemically with judgements of truth". Paradis (2003, p. 214) clarifies that "such pragmatic conditions act as motivating forces on the conceptual representation evoked by *really*". The epistemic meaning has conceptual underpinnings: truth attesting of the proposition, subjective emphasis of the situation and reinforcement of scalar properties. In truth attesting "*really*", the evidence reflects the reality concept evoked by "*actually*" whereas in emphasising "*really*", the evidence of truth is indirect via subjective emphasis made by the speaker. Lastly, "in the case of *really* as a reinforcer, the evidence of truth conveyed is indirect through *really* as a degree operator" (Paradis, 2003, p. 216).

Below is a table summarising the distribution of the occurrences of "*really*" in the Brazilian learner corpus. Altogether, there were 57 occurrences of "*really*". Out of these, 10 occurrences were disregarded as they were embedded in the interviewer's fixed rubrics. The final relative frequency for 47 occurrences was 19.7454 per 10,000 words.

Table 4: Functions of "*really*" in the Brazilian learner corpus:

Really	**Raw frequencies**	**Percentage**
Degree reinforcer before scalar adjectives	20	42.55%
Emphasiser adjacently placed to attitudinal verbs	20	42.55%
Truth attester utterance initial position utterance final position	4	8.52%
De-emphasiser preceding negation	3	6.38%
Total occurrences	47	100.00%

The Brazilian data indicate that the subjects of this study used "*really*" mostly as a degree reinforcer of scalar adjectives and as an emphasiser accompanying attitudinal verbs. The above figures for their occurrences were drawn upon the formulae "*really + adjective*" and "*really + verb*". All scalar adjectives collocating with "*really*" showed a positive connotation in their actual realisation: important (4)[46], nice (4), good (3), different (2), popular (1), smart (1), interesting (1), impressed (1) and close (1). "*Really" + "fun*" (1) and "*really" + "quickly*" (1) were also categorised within the degree reinforcing category. The attitudinal verbs collocating with "*really*" were as follows: like (9), enjoy (6), need (1), love (1), know (1), wanted (1) and would like (1- projected).

If we compare the above data with Paradis's native speaker findings, one can infer that the Brazilian learners of this study were able to produce "*really*" so as to perform all the three pragmatic functions proposed by Paradis (2003), including the de-emphasiser one, though overall relative frequencies differed. Both Brazilian learners and native speakers produced "*really*" in a similar order: degree reinforcer, emphasiser and truth attester. Within the distribution among functions, the Brazilian learners displayed a higher percentage of

[46] The number in brackets refers to the number of occurrences.

emphasising "*really*" and a lower one of truth attesting if compared to native speaker data. This difference may signal that the Brazilian learners of this study tend to produce "*really*" more frequently within utterances, either to emphasise scalar properties of adjectives or to reinforce the subjectivity of situations when accompanying attitudinal verbs. In the Brazilian learner data, there were 3 occurrences of "*really*" in utterance initial position and 1 occurrence in utterance final, possibly indicating less familiarity with "*really*" truth attester, which takes scope over full propositions.

Let us now examine some occurrences of "*really*" in the Brazilian corpus. In the conversation below (10), we can identify one occurrence of "*really*" as a degree reinforcer:

(10)

<S1> Some people say that going out to relax is a waste of time and money. Do you agree?
<S5> Not at all, I don't think. I think it's important uh after a <SE> pause</SE> difficult day at work or a <SE> pause</SE> busy week, it's important to do things to relax and to take your mind off the problems. I think it's **really** important.
(FCE P4- B1 BR-02)

In the above conversation (10), the interviewer presents a controversial statement: "Some people say that going out to relax is a waste of time and money". Subject 3 then marks her strong disagreement with the previous proposition through the use of the expressions "*Not at all, I don't think*". She then indicates reasons for going out to relax: it's important to relax after a difficult day or a busy week, it's important to do things to relax, it's important to take your mind off the problems. The use of "*really*" in the last utterance emphasises the degree of importance of going out to relax.

(11)

<S16> Yeah, good, good thing because the people want to see the soccer's game, for example, together with another persons so it's good because you don't want to stay at home and saw it alone <SE> laughing </SE> so you can go to a café and seeing the TV there. I think this is **really** important too, the menu, international menu cause there can, everyone can read and understand everything and order, I think they **really** know what they are order and not guessing <SE> laughing </SE> what is.
(FCE P3- B1 BR-08)

In (11), subjects are discussing suggestions to improve the popularity of a café. Two examples of "*really*" are produced within an individual long turn. In the utterance "*I think this is **really** important too, the menu…", "this"* is used cataphorically as it refers to the menu, which appears later in the text. *"Really*" is therefore used as a reinforcer of the degree of importance of the menu. In the second occurrence "*I think they **really** know*", "*really*" precedes the state verb "*know*" and it is used as an emphasiser.

(12)

<S1> Tell us about a day you've really enjoyed recently.

…

<S16> A day, uh on Sunday I went to the game, to the match in Rio, in Brazil, and stay <SE> laughing /SE>.

<S1> Oh nice!

<S16> <SE> laughing /SE> And with nice result because Brazil win, yeah, é but it's a **really** nice game, uh have many Brazilians <SE> laughing /SE> people and it's **really** nice having <SE> pause /SE> they saw and they heard everything uh it's **really** interesting.

(FCE P1- B1 BR-08)

The above conversation (12) displays three instances of "*really*" produced by the same subject as in conversation (11). A repair sequence has been omitted. Despite the noticeable problems with verb tense consistency, among other grammar problems, subject 16 manages to sound slightly fluent in her long turn through the lexical repetition of "*really*". In the two collocations with "*nice*", "*really*" is used as a reinforcer of the degree of pleasantness of the football match attended by subject 16. In the final utterance of the long turn, "*it's **really** interesting*" seems to have been produced as an afterthought, summarising the ideas previously expressed in the turn. The use of "*really*" reinforces how interesting the match was.

Furthermore, emphasising "*really*" was also found preceding two unfinished utterances:

(13)

<S1> …, where would you like to go on holiday in the future?

<S5> In the future, uh to Paris.

<S1> Why?

<S5> Because I love everything about Paris and hum I **really**, it's the first place that comes to mind that I wanna know.
(FCE P1- B1 BR-02)

(14)
<S1> Uh … tell us about a day you've really enjoyed recently.
<S10> A day that I **really** <SE> pause</SE>+
<S1> <SE> overlapping </SE>Enjoyed.
<S10> + enjoyed? Hum, let me see <SE> pause</SE>.
(FCE P1- B1 BR-02)

In (13), the linguistic co-text indicates the use of "*really*" as an emphasiser as in "I would really like to go to Paris". In (14), there is a self-initiated repair sequence in which subject 10 simply repeats part of the previous rubric until "*really*" and then signals that he requires clarification. The rubric produced by the interviewer displays "*really*" also as an emphasiser followed by the attitudinal verb "*enjoy*".

In relation to "*really*" as a de-emphasiser, different structure types were found:
(15)
<S13> Remember that? I guess it's great. And <SE> pause</SE> I don't know, I don't like too, too much about music when it's too loud in a <SE> pause</SE> in a café or in a venue or something like this.
<S12> <SE> overlapping</SE> To pass the point.
<S13> And I am a musician but I **really** don't like that in a café because they are too <SE> pause</SE>
<S12> <SE> overlapping</SE> I think if you stay +
<S13> near to you.
(FCE P3- B1 BR-06)

In conversation (15), subjects have just discussed the advantages of offering a free coffee and of having comfortable seats as ways to attract more customers to a café. Subject 13 has given an example of a café that offers these items. He then launches the idea of having live music. He expresses a negative attitude towards it through the negative forms of the state verbs: "*I*

don't *know, I* ***don't*** *like too, too much about…"*. In his subsequent turn, "*really*" is used as a de-emphasiser, attenuating the fact that despite being a musician, he does not like loud music. The overlapping turn produced by subject 12 possibly characterises language transfer. In Brazilian Portuguese, the idiomatic expression "*passar do ponto*" means that something stops being good when in excess. Thus, it seems that the wording of the expression was literally translated into English.

As mentioned by Paradis (2003, p. 214), degree reinforcer "*really*" only occurs in statements. In negative sentences, it becomes a de-emphasiser as it has no tone. Thus, despite being followed by a scalar adjective (good), "*really*" in (16) fits the characterisation of a de-emphasiser:

(16)

<S20> Uh if it is open in the night<SE> pause </SE> I don't know, I, I think it's not **really** +

<S21> <SE> overlapping </SE> Yeah, not sure if it's a good idea also.

<S20>+ good because uh a coffee maybe if, if it sells more than coffee but just coffee at night, I don't know if+

<S21> <SE> overlapping </SE> Yeah +

<S20> +if it would be a good, good idea.

(FCE P3- B1 BR-10)

Finally, let us introduce some examples of truth attesting "*really*":

(17)

<S10> + uh an expresso. And I <SE> pause</SE> but I also like to order a café and go outside, in the winter.

<S11> **Really**?

<S10> **Really**, **really**. It's like a New York way of life. Go to the subway with a café and uh maybe with a kindle +

<S11> Uhuh.

<S10> + in the <SE> pause</SE> other hand.

<S11> </SE> nice.

<S10> **Really** something.

(FCE P3- B1 BR-06)

In the above conversation (17), there are four examples of "*really*" as a truth attester. Initially, subject 10 makes an assertion that is somehow surprising for subject 11: He likes to order a coffee and then drink it outside in winter. Subject 11 shows his disbelief through the use of "*really*" in the interrogative form as a truth attester. In the following turn, subject 10 produces two consecutive examples of "*really*", attesting the truth of his previous assertion. The last occurrence of "*really*" in "*really something*" is in a tonic position and it is also used as a truth attester of a previous proposition, contrasting to what subject 11 may think about it.

The above analyses of the Brazilian learner corpus illustrate the different uses of "*really*" as produced by the subjects of this study so as to convey pragmatic meaning. While learners produced "*really*" mostly before scalar adjectives as a degree emphasiser and adjacently placed to attitudinal verbs as a reinforcer, there were some occasional uses of truth attester "*really*" in initial and final utterance positions. As for the distribution of "*really*" among interviews, 8 out of 10 interviews included "*really*": interview 1 (2 occurrences), interview 2 (3 occurrences), interview 3 (1 occurrences), interview 5 (4 occurrences), interview 6 (3 occurrences); interview 8 (21 occurrences), interview 9 (1 occurrence), interview 10 (12 occurrences). A higher frequency of "*really*" was observed in interviews 8 and 12, which may indicate individual variability.

For comparative purposes, below is a table indicating the raw and the relative frequencies (per 10,000 words) of "*really*" in different positions in the Brazilian learner corpus and in the Spanish component of LINDSEI (112,921 words spread in 50 texts):

Table 5: Position distribution of "*really*" in 2 learner corpora

	Brazilian corpus		**Spanish corpus**	
Really	Raw frequencies	Relative frequency	Raw frequencies	Relative frequency
Utterance-initial	3	1.2603	2	0.1771
Utterance - final	1	0.4201	14	1.2398
Other positions	43	18.0649	329	29.1354
Total	**47**	**19.7454**	**345**	**30.5523**

The above data indicate that, by large, both Brazilian and Spanish learners produced "*really*" inserted within utterances. As for the occasional occurrences of "*really*" in initial and final positions in both corpora, the Brazilian learners produced "*really*" more frequently in initial position while the Spanish learners displayed a firmer command of "*really*" in final position. Overall, the Spanish learners produced "*really*" with a much higher frequency than that of the subjects of this study, even higher than that of native speakers (COLT: 26.96 per 10,000).

For illustrative purposes, concordance lines of "*really*"+ "*I*" retrieved from the face-to-face conversations from the 1990s sub-corpus have been included so as to illustrate different meanings of "*really*".

absolutely <laugh> absolutely uh But that 's our our bread and butter	really I	mean in terms of the way we work we 're constantly confronted
did you decide at that stage to continue in architecture No not	really I	kind of , > I mean , > I I
It was the course It was a kind of combination of things	really I	did n't really know the uni the the sort of scene at
perhaps , , > having sex with your father is n't	really i	e that 's what you were feeling which is why ,
I did n't like uh , > was leaving my mum	really I	think that was it I was I was insecure Uhm ,
, > it was so long uh at primary school I	really I	did n't like it why I pretended to be ill That was
you could tell Did she succeed No No She did n't succeed	really I	mean like you knew that uh although we like we we me
> And how 's work going now That 's fine now I	really I	hated London last year but uhm You got used to it Yes
you know , > I do n't do anything with it	really I	just sort of word process really , > But someone who
, > It is the only way to grow them yourself	really I	mean and plant them out , , > What you want

Figure 5: Extract of concordance lines for "*really*" + "*I*" in the face-to-face conversations from the 1990s sub-corpus

The above concordance lines display a number of examples of "*really*" truth attester. In "*that's our our bread and butter* ***really****, I mean…*", "*It was a kind of combination of things* ***really***" and "*It's the only way to grow them yourself* ***really****, I mean*", "*really*" is placed at the end of the propositions, taking scope over them. In addition, the examples in which "*really*" is followed by "*I mean*" indicate a further refinement or justification of the previous assertions. The examples "*She* ***didn't*** *succeed* ***really***" and "*I* ***don't*** *do* ***anything*** *with it* ***really***" show propositions with negative elements. As "*really*" is placed at the end of the propositions, it takes scope over them, functioning as a truth attester too. In the example "*having sex with your father* ***isn't really****…*", "*really*" is in medial position and it is preceded by a negation, working as a de-emphasiser.

To conclude, in this section we have investigated the pragmatic functions of "*really*" and presented native speaker data and Brazilian and Spanish learner data. Paradis (2003)

categorises "*really*" as a truth attester, an emphasiser and a degree reinforcer. By comparing COLT and Brazilian learner data, one can infer that the subjects of this study were able to produce "*really*" so as to perform all the three pragmatic functions proposed by Paradis (2003). Within the distribution among functions, the Brazilian learners displayed a higher percentage of emphasising "*really*" and a lower one of truth attesting "*really*" if compared to native speaker data. This difference may signal that the Brazilian learners of this study tend to produce "*really*" more frequently within utterances. In the Brazilian learner data, there were 3 occurrences of "*really*" in utterance initial position and 1 occurrence in utterance final, possibly indicating less familiarity with "*really*" truth attester. As for the distribution of "*really*" among interviews, 8 out of 10 interviews included "*really*". Comparisons with other learner data (LINDSEI) indicated that both Brazilian and Spanish learners produced "*really*" mostly inserted within utterances. Overall, the Spanish learners produced "*really*" with a much higher frequency than that of the subjects of this study.

4.1.3. "*Actually*"

This section initially details Aijmer`s (2002) categorisation of the functions of "*actually*" and her corpus-based findings vis-à-vis native speaker spoken production (London-Lund). It then presents Brazilian learner corpora data (position, frequency and functions), qualitative analyses and comparisons with native speaker and Spanish learner data (LINDSEI). It also includes concordance lines (BNC Sampler spoken sub-corpus) so as to illustrate less frequently observed functions or collocations in both learner data.

Aijmer (2002) investigates the uses of "*actually*" in two spoken corpora: The London- Lund Corpus (LLC) and the Corpus of London teenager speech (COLT). Aijmer (2002) points out that utterances usually convey speakers' attitude to the message. When speakers produce factual propositions and are concerned about the truth conditions of their utterances, they rely on evidentials to convey different attitudes towards knowledge and truth. "*Oddly enough*" is a category of this type as it signals a conflict between what speakers had expected and what is known. "*Of course*" shows that something is in line with expectations. "*In fact*", "*actually*", "*really*", "*in actual fact*" and "*as a matter of fact*" signal that an event is surprising or unexpected.

"*Actually*" is formally an adverb that has also developed into a discourse particle through grammaticalisation. In its basic meaning, it expresses a discrepancy between reality and what appears to be the case. Similarly to "*sort of*" and "*like*", "*actually*" is an adjuster word by the use of which another word is "adjusted to meet the innumerable and unforeseeable demands of the world upon language" (Austin, 1962: 73). However, when it is used as a discourse particle, "*actually*" allows speakers to change perspective and to call attention to something they have just come to think. Let us examine these differences in meaning in the two examples provided by Aijmer (2002, pp. 252-253):

(1)

"*As I understand this thing, there is deep opposition to this, from heads of departments, and from the departments themselves and from other things which aren't* ***actually*** *academic departments, there is deep skepticism about it*".

(2)

"*One I remember best was lemon sole, no, paté du maison, la maison, followed by lemon sole, followed by steak, followed by cream, including wine and tip, for fifteen francs, which had five courses.* ***Actually*** *I'm quite sure there was bits of cheese as well".*

In the first example, "*actually*" conveys its etymological meaning (in actuality) as the speaker draws a distinction between academic departments and things that are not academic departments, since the latter differ from prototypical departments. Thus, "*actually*" is used as a regular adverb. In the second example, "*actually*" is used to mark the addition of detailed information and, therefore, calls attention to something new, functioning as a discourse particle. Nonetheless, Aijmer (2002) clarifies that this distinction has been disputed as "*actually*" does not have a very precise meaning. She examines other criteria to distinguish between the propositional use of "*actually*" and the discourse particle.

Utterance position plays a part in the distinction between the adverb and the discourse particle.When "*actually*" is placed medially, it hedges the proposition or a single element. When it is placed initially or finally, it is a discourse participle with textual or interpersonal function. "*Actually*" can appear in different positions depending on whether it is being used as an adverb or as a discourse particle (Aijmer, 2002, pp. 256-259). In pre-verbal position as an adverb, it is an emphasiser with a reinforcing effect on the truth value or part of the clause it

belongs to, e.g. "*He* **actually** *said something totally different*". When "*actually*" is an emphasising adverb, it can be positioned before or after the verb "*to be*", e.g. "…*called Wilfred Evans, who is* ***actually*** *a very nice chap*". In addition, it is usually placed after a negation, e.g. "…*and I didn't* ***actually*** *speak to her*". As a discourse particle, "*actually*" in initial position plays the role for the development of the discourse and thus, have a cohesive function, e.g. "***Actually*** *he caught me on the hop because…*". When it is placed in final position, the particle functions as a floor-holder, emphasising the relationship between the speaker and the hearer, for instance (Aijmer, 2002, p. 258):

(3)

A: "*Yes, people didn't think you were going to be in today,* ***actually***".

B: "*Well, they never think I'm in today* ***actually***."

"*Actually*" can be placed in a clause final position, which is not necessarily an utterance final position, e.g. "*I think it would help a lot,* ***actually,*** *if we would, we could wait*". "*Actually*" in utterance-final and clause-final position appears to perform a social function. In a post-head position, it is more like an intensifier, e.g. "*we have made quite a round trip,* ***actually****, in the last three years*".

Aijmer (2002) investigates the frequency distribution of "*actually*" both as an adverb and as a discourse particle over different text types in the London-Lund Corpus of Spoken English (LLC). Table 6 (Aijmer, 2002, p. 259) displays text types, raw frequencies and relative frequencies (per 10,000 words).

Table 6: Distribution of "*actually*" over different text types in LLC

Text type	**Raw frequencies**	**Relative frequencies**
Face surreptitious	271	12.4
Face non-surreptitious	77	12.8
Telephone	50	8.3
Discussion	67	6.7
Prepared	3	0.75
Public	12	1.8
Total	**426**	

Aijmer's (2002) results indicate a higher frequency of "*actually*" in face-to-face conversations, both in conversations recorded surreptitiously and non-surreptitiously, followed by telephone conversations and discussions. The lower frequencies of "*actually*" in prepared and public speeches seem to signal a stronger association of "*actually*" with less formal types of conversation. Aijmer (2002) also investigates "*actually*" specifically as a discourse participle in initial and final (utterance-final, clause-final or post-head) positions as shown in table 7 (per 10,000 words).

Table 7: Position distribution of "*actually*" over different text types

Actually	**Utterance initial**	**Medial position**	**Final position**
Face surreptitious	32	89	96
Face non-surreptitious	10	28	39
Telephone	5	18	27
Discussion	9	43	15
Public	1	8	3
Prepared	X	3	
Total	**57**	**189**	**180**

Aijmer's findings (2002, p. 260) indicate that the most frequent use of "*actually*" is in medial position, being employed as an adverb. Conversely, the most typical use of the discourse particle is in final position, being three times more frequent than in initial position. "*Actually*" in final position seems to suggest its role as a style enhancer, due to its high frequency in informal conversations.

Aijmer (2002) also discusses the pragmatic functions of the discourse particle "*actually*". One of the main functions of "*actually*" (as "*in fact*") is to indicate a contrast with the preceding utterance. "*Actually*" allows speakers to distance themselves from the factuality of a previous utterance and to express a contrast with it. The contrastive "*actually*" expresses an opposition between different points of view; it may be used to deny an implicit belief, to signal a counter-claim or an objection, to facilitate for speakers to take up a new position, to switch to another aspect of the topic or to reformulate an utterance. It may also be used to mark a correction. When it co-occurs with "*no*", it marks a counter-claim, a correction or incipient disagreement with a preceding speaker. "*Well actually*" introduces a clarification of what has just been said. When it occurs in the preface of a face-threatening act, it sounds apologetic and disarming. Aijmer (2002) argues that despite its contrastive function, "*actually*" conveys politeness.

"*Actually*" also has an empathic function when used to signal an explanation or justification. This emphasis may be further strengthened by accompanying expressions such as "*to tell you the honest truth*". "*Actually, I think*" can be used to reinforce speakers' point of view expressed by other individuals. Another function of "*actually*" is to indicate that a preceding move needs to be elaborated or exemplified. It may also indicate that some amendment (correction, clarification) needs to be made before the conversation can continue. In such cases, "*I mean*" and "*actually*" may collocate. "*Actually*" in its interpersonal function serves as a style-marker and as a politeness marker. When it is placed in end-position, its contrastive function is weak as it is oriented to the speaker-hearer relationship as a softener. It may also function to foreground a subjective opinion or as an afterthought ("*come to think of it*"). And when it co-occurs with a negation, it indicates that the speaker is defensive.

Now that we have detailed the distribution and the functions of "*actually*" in a native speaker corpus (LLC), let us focus on learner data. In the Brazilian learner corpus, there were 4 occurrences of *"actually"* produced by 3 subjects in 3 different interviews. Table 8 summarises raw frequencies, relative frequencies (per 10,000 words) vis-à-vis utterance position.

Table 8: Position distribution of "*actually*" in the Brazilian learner corpus

	Brazilian corpus	
Actually	Raw frequencies	Relative frequency
Utterance-initial (also preceded by filled pauses)	3	1.2603
Medial	1	0.4201
Utterance final	x	x
Total	**4**	**1.6804**

The above results indicate that "*actually*" was mostly used as a discourse particle since 3 out of 4 occurrences were placed in initial position. All 4 occurrences were followed by "*I*" and verbs expressing preferences. In the examples of the discourse particle, "*actually*" was used to introduce one's opinion or preference conveying different attitudes to the preceding discourse.

The single occurrence of "*actually*" in medial position displayed its use as a mitigating device. The low frequencies of "*actually*" in the Brazilian data may indicate learners' underuse of the discourse marker and of the regular adverb. Furthermore, the absence of "*actually*" in final position may signal lack of familiarity with its functions in this position, such as a floor holder, a marker of subjective opinion or an afterthought.

Let us now examine each of the above occurrences in context. In the conversation below (4), subjects are discussing whether or not they would like to work in a café. The extract displays 1 occurrence of "*well*" and 1 occurrence of "*actually*". "*Well*" is placed in a mid-utterance break and used in its change-related function (repair). It is interesting to notice that it takes subject 20 three turns to eventually state his opinion. His previous turns are marked by filled and unfilled pauses. "*Actually*" is used to introduce and to soften his opinion as it constitutes a dispreferred response to the question proposed by Subject 1.

(4)

<S1> … would you like to spend time in a café like this?

<S20> Uh <SE> pause </SE>

<S21> Ah, not too much time but +

<S20> <SE> overlapping</SE> <SE> laughing </SE>

<S21> + **well**, sometimes it's nice to take a coffee <SE> pause </SE>+

<S20> <SE> overlapping </SE> Yes, have <SE> pause </SE>

<S21>+ in the middle of your activities there but <SE> pause </SE>

<S20> **Actually** I don't like coffee, I don't like the taste of coffee, I don't drink it +

<S21> <SE> overlapping</SE> Yeah +

<S20><SE> overlapping</SE> + drink it so much

(FCE P4- B1 BR-10)

In the example below (5), "*actually*" is in initial position following a filled pause "*uh*". It also has a discourse marking function as subject 13 gives detailed information about his preferences:

(5)

<S1> Uh which type of music would you prefer to listen to?

<S13> **Uh actually** I like all of kinds of music but mostly I like uh rock and blue and <SE> pause</SE>+

<S1> <SE> coughing</SE> Sorry.
<S13> + hard rock, things like that.
(FCE P2- B1 BR-06)

Similarly to (5), conversation (6) displays "*actually*" in initial position, preceded by "*uh*". It has a discourse function since it introduces information that may be somewhat unexpected: subject 20 prefers to spend time on this own rather than with other people.

(6)
<S1> … do you prefer to spend time on your own or with other people?
<S20> **Uh actually** I prefer spend my time on my own <SE> pause</SE> +
<S1> <SE> overlapping</SE> Why?
<S20>+ I prefer read a book, study a little bit uh, I don't know I really enjoy uh be my own in, and no, and no with other people, I don't know, I think it's a kind of, of thing like that dentists likes <SE> laughing </SE> they are very , very <SE> pause</SE> uh <SE> pause</SE> they prefer be, be to uh alone.
(FCE P1- B1 BR-10)

In the conversation below (7), "*actually*" is preceded by a filled pause and placed between two occurrences of the pronoun "*I*", thus, in medial position. Subjects are discussing suggestions to improve the popularity of a local café, based on pictorial prompts. One of the prompts shows a television set broadcasting a football match. Subject 14 has the assumption that the primary role of a television set in a café would be to show football matches. She uses "*actually*" to politely disagree with her peer's suggestion and to introduce a justification why a television set would not be a good idea to improve the popularity of a café. In this case, "*actually*" is used as a hedge, mitigating her assertion.

(7)
<S15> And how about the television? What do you think about?
<S14> **Yes, uh I actually I don't like** football +
<S15> <SE> overlapping/SE> I do.
<S14> and I think that a lot of uh, uh , you go to a coffee to, to relax and spend time with your friends, talking, not watching TV or some games so.
<S15> Yeah, I, I like football but uh I don't think that uh televisions uh <SE> pause/SE> will make your coffee more successful.

<S14> Yeah, it's, it's similar that you spend time in coffee using a Internet.
< S15> Yeah.
(FCE P3- B1 BR-07)

Now that we have presented the functions of "*actually*" in the Brazilian corpus, let us compare results with other learner data (Spanish component of LINDSEI). Table 9 displays the results of both Brazilian and Spanish learner data vis-à-vis position distribution (per 10,000 words).

Table 9: Comparison between the position distribution of "*actually*" in 2 learner corpora

	Brazilian corpus			**Spanish corpus**
Actually	Raw frequencies	Relative frequency	Raw frequencies	Relative frequency
Utterance-initial (also preceded by filled pauses)	3	1.2603	5	0.4427
Medial	1	0.4201	33	2.9223
Utterance final	x	x	3	0.2656
Total	**4**	**1.6804**	41	**3.6308**

Overall, the above results indicate that the Spanish learners produced "*actually*" with a higher frequency than that of the subjects of this study. They were also able to place "*actually*" in all utterance positions, including utterance final, performing discourse particle and regular adverb functions. This may be due to the fact that LINDSEI learners are at a higher level of language proficiency (as from CEFR B2). Having said that, most occurrences of "*actually*" were placed in medial postion, signalling a tendency by the Spanish learners to use it mostly as an adverb. In addition, the Spanish learners produced "*actually*" initial with a lower relative frequency than that of the subjects of this study.

As for the frequencies of "*actually*" in the benchmark corpora, the BNC Sampler spoken sub-corpus displayed 1,177 occurrences while 515 occurrences were found in the specialised face-to-face conversations from the 1990s. Relative frequencies were normalised to 10,000 words amounting to 10.3240 and 18.9784, respectively. By comparing relative frequencies, one can

notice a much higher frequency of "*actually*" in the specialised sub-corpus, aligned with Aijmer's (2002) findings that "*actually*" is primarily used in face-to-face conversations. When one compares the frequencies between native speaker and learner corpora, the underuse of "*actually*" by both Brazilian and Spanish learners becomes more apparent.

Furthermore, after conducting standard queries of "*actually*" in the benchmark corpora, I used the CQPweb tool "frequency breakdown of words only" so as to investigate the most frequent collocating items with "*actually*", 1 position to the right. In the BNC Sampler spoken sub-corpus, the four most frequent collocations (case-insensitive) were as follows: "*actually*" followed by commas (85 hits), full stops (76 hits), "*got*" (36 hits) and "*I*" (27), respectively. Likewise, the face-to-face spoken conversations from the 1990s sub-corpus displayed the following most frequent collocations: punctuation (either commas or full stops – 34 hits), "*I*" (32 hits), "*it*" (12 hits) and "*the*" and "*you*" (both with 10 hits each). While the BNC Sampler spoken sub-corpus seems to indicate a tendency of "*actually*" in clause final position (both commas and full stops), the face-to-face conversations from the 1990s shows a similar distribution between clause final and "*actually I*". All occurrences in the Brazilian learner corpus displayed "*actually*" followed by "*I*", in contexts of opinions or preferences.

Let us illustrate some examples of "*actually*" in utterance final position, taken from the BNC Sampler spoken sub-corpus as shown in figure 6:

have anything Catherine on that tonight ? . Yes yes we do	actually	. Mm , mm . Erm , a couple of days ago
things like that I ne , hope and anticipate that we can	actually	. I can only guess that for erm some reason the bolts
to the the weaknesses which are connected to some of the threats	actually	. Right okay we 'll work from the top down actually .
threats actually . Right okay we 'll work from the top down	actually	. Alan there quite a big weakness within the company . Erm
you is do n't abandon hope erm sounds like a religious text	actually	. Do n't abandon hope just because you think oh dear ,
er Malcolm and Liz . . The Liz is just arriving ,	actually	. Is she ? Oh good . Yes , they were parking
sort of erm . It 'd be cheaper to buy it ,	actually	. progressing , are n't they ? No . is it really
the end is Hark the Herald is a good way to end	actually	. Hark the Herald is an excellent ending . Yes , .
a a complete contrast . Erm it could n't be more different	actually	. Er you know living er so close to neighbours and people
what what impact does that have on the actual problems that they	actually	. To what ex what I 'm saying is , To what

Figure 6: Extract of concordance lines for "*actually*" + "." in the BNC Sampler spoken sub-corpus

The instances of *"actually"* followed by a full stop display its use as a discourse particle in both clause and utterance final positions. In "*Yes, yes we do* ***actually***", it is used as a softener, mitigating the assertion. In "*Hark the Heralld is a good way to end* ***actually***", it foregrounds a subjective opinion. "*Right okay we'll work from the top down* ***actually***", it seems to have been used as an afterthought, since the first words of the utterance "*Right, okay we'll…*" convey a spontaneous decision. In "*The Liz is just arriving,* ***actually***", it appears to have been used as a floor holding device.

In this section we have investigated the discourse and the pragmatic functions of "*actually*" and presented native speaker and learner data. Aijmer (2002) draws a distinction between the regular adverb and the discourse particle based on utterance position. When "*actually*" is placed medially, it hedges the proposition or a single element. When it is placed initially or finally, it is a discourse participle with textual or interpersonal function. Aijmer's (2002) findings in the LLC corpus indicate that the most typical use of the discourse particle is in final position, being three times more frequent than in initial position. The Brazilian learner corpus displayed 4 occurrences of "*actually*", 3 of which were in initial position performing discourse functions while 1 occurrence was placed in medial position and employed as a mitigating device. Comparisons with benchmark frequencies signalled Brazilian and Spanish learners' underuse of "*actually*", both as a discourse marker and as a regular adverb.

4.2. DISCUSSION AND RECOMMENDATIONS

So far we have presented corpus-based findings vis-à-vis the pragmatic functions and relative frequencies of the discourse marking adverbs "*well*", "*really*" and "*actually*", based on British English native speaker spoken data and Brazilian and Spanish learner data. Let us now discuss these findings in the light of the specific research sub-question addressed in this chapter.

RSQ1: *What are the most common discourse marking adverbs used to mediate segments of discourse in conversations by Brazilian learners at CEFR B1? How do their frequency of use and pragmatic functions compare to those of native speakers?*

Let us initially focus on the discourse marking adverbs present in learners' oral production. As shown in table 1 (section 4.1), the Brazilian learners of this study produced the following

most frequent adverbs that seem to fit the characterisaton of discourse markers: "*maybe*", "*really*", "*just*", "*well*", "*probably*" and "*actually*". The adverbs "*really*", *"well*" and "*actually*" were chosen for investigation in chapter 4 since they are commonly used to mediate segments of discourse with regards to the different parts of discourse and also to the relations between speakers and parts of discourse. The adverbs "*maybe*", "*just*" and "*probably*" are traditionally associated with mitigation and were chosen to be investigated in chapter 5.

The range of adverbs produced by the Brazilian learners attests a limited command of discourse marking adverbs in general. Table 10 displays the raw frequencies of the specific discourse marking adverbs under investigation as produced by individual learners.

Table 10: Distribution of discourse marking adverbs per subject

Brazilian learners	Well	Really	Actually
Subject 2			
Subject 3		2	
Subject 4			
Subject 5		3	
Subject 6			
Subject 7		1	
Subject 8			
Subject 9			
Subject 10		2	
Subject 11	5	2	
Subject 12		1	
Subject 13		2	1
Subject 14			1
Subject 15			
Subject 16		20	
Subject 17		1	
Subject 18	3	1	
Subject 19			
Subject 20		7	2
Subject 21	1	5	
Total	**9**	**47**	**4**

The above results demonstrate that 65% of the subjects of this study produced at least one form of discourse marking adverbs used to mediate segments of discourse. While "*really*"

was used by 60% of the subjects, "*well*" and "*actually*" were employed by only 15% of the subjects. In addition, 25% of the subjects produced two forms of discourse marking adverbs, showing a more consistent command. One subject of this study (subject 16) overused "*really*". 35% of the subjects did not produce any discourse marking adverb.

Let us now summarise results related to specific forms. The Brazilian learner corpus displayed 9 occurrences of "*well*" as a one-word pragmatic marker distributed in 3 interviews. The comparison of relative frequencies (per 10,000 words) between the native speaker benchmark corpora and the Brazilian learner corpus confirmed considerable underuse of the discourse marker "*well*" by the subjects of this study:

a) BNC Sampler spoken sub-corpus: 54.3833

b) Face-to-face conversations from the 1990s: 63.1264

c) Brazilian learner corpus: 3.7810

Furthermore, the subjects of this study used "*well*" to perform the following functions as shown in table 3 (section 4.1.1):

a) Speech management functions (66.7%): choice related 44.5%, change related 22.2%

b) Attitudinal functions: opinion 33.3%

Based on the above data, it is possible to conclude that all occurrences of "*well*" portrayed its use as a discourse marker, mediating segments of discourse: choice related (for pausing and planning what to say next), change related (for repair) and opinion (for conveying stance). With regards to utterance position, instances of "*well*" occurred in three contexts: after filled and unfilled pauses (utterance initial), in utterance initial and in medial positions. "*Well*" was mostly used for speech management functions and often preceded by filled and unfilled pauses in initial position or in medial position, being used as a floor holding device, similarly to Aijmer's results (2011).

Native speaker data (LOCNEC) indicated that following distribution among functions (Aijmer, 2011, p. 248):

a) Speech management functions (65%): choice related 6.4%, change related 26.1%, prospective 20.2%, marking stages 3%, quotative 9.3%

b) Attitudinal function (35%): opinion 12.3%, disagreement 22.7%

Comparisons between Brazilian learner and native speaker data (LOCNEC) demonstrated that both used "*well*" more frequently for speech management functions. However, native speakers used "*well*" in the change-related function more frequently than Brazilian learners did while the latter used "*well*" in the choice-related function with a higher percentage among functions than native speakers did. Percentages for attitudinal "*well*" were also higher in native speaker data.

Secondly, the subjects of this study produced 47 occurrences of "*really*" distributed in 8 interviews. However, 20 instances were produced by the same subject.The comparison of relative frequencies (per 10,000 words) between the native speaker benchmark corpora and the Brazilian learner corpus indicated a consistent use of "*really*" by the Brazilian learners, with a higher frequency than that in the BNC Sampler and a lower one than that in the specialised face-to-face sub-corpus.

a) BNC Sampler spoken sub-corpus: 15.3501
b) Face-to-face conversations from the 1990s: 29.0389
c) Brazilian learner corpus: 19.7454

The Brazilian learner corpus displayed the following distribution of "*really*"among functions as shown in table 4 (section 4.1.2):

a) Degree reinforcer (before scalar adjectives): 42.55%
b) Emphasiser (adjacently placed to attitudinal verbs): 42.55%
c) Truth attester (utterance initial and final): 8.52%
d) De-emphasiser: 6.38%

The above data allow us to conclude that the Brazilian learners of this study were able to employ "*really*" to perform all the pragmatic categories proposed by Paradis (2003). The subjects of this study used "*really*" more frequently as a pragmatic marker modifying speaker meaning, considering the degree reinforcer, emphasiser and de-emphasiser functions. While the degree reinfoncer "*really*" reinforces scalar properties denoted by adjectives, the emphasiser "*really*" has the effect of adding subjective emphasis of situations and accompanies attitudinal verbs. As for discourse properties, truth attester "*really*" has an epistemic function, since it ensures the truth of the assertion that it takes in its scope and

provides implicit evidence based in reality. Truth attester "*really*" was observed in utterance initial and final positions mediating segments of discourse, but with a lower frequency.

Native speaker data (COLT) indicated the following distribution among functions (Paradis, 2003, p. 197):
a) Truth attester: 23%
b) Emphasiser: 33%
c) Degree reinforcer: 44%

Comparisons between native speaker data (COLT) and Brazilian learner data showed that both used "*really*" mostly as a pragmatic marker modifying speaker meaning within utterances, either to emphasise scalar properties of adjectives or to reinforce the subjectivity of situations when accompanying attitudinal verbs. Native speakers produced truth attester "*really*" more frequently than Brazilian learners did, which suggests that the subjects of this study are less familiar with this discourse function of "*really*".

Thirdly, the Brazilian learner corpus displayed 4 occurrences of "*actually*" distributed in 3 interviews. The comparison of relative frequencies (per 10,000 words) between the native speaker benchmark corpora and the Brazilian learner corpus indicated underuse of the discourse marker "*actually*" by the subjects of this study:
a) BNC Sampler spoken sub-corpus: 10.324
b) Face-to-face conversations from the 1990s: 18.9784
c) Brazilian learner corpus: 1.6804

In relation to functions, the Brazilian learners used "*actually*" mostly as a discourse marker. "*Actually*" was placed in initial position in 3 occurrences and used to introduce one's opinion, conveying different attitude to the preceding discourse, i.e. mediating segments of discourse. The single occurrence of "*actually*" in medial position displayed its use as a mitigating device, being used as a pragmatic marker. All 4 occurrences were followed by "*I*" and verbs expressing preferences.
Native speaker data (LLC) indicated the following distribution among utterance positions (Aijmer, 2002):
a) Utterance initial: 13.38%

b) Medial: 44.37%
c) Utterance final: 42.25%

Aijmer's native speaker findings (2002, p. 260) show that the most frequent use of "*actually*" is in medial position, being employed as an adverb. Conversely, the most typical use of the discourse particle is in final position, being three times more frequent than in initial position.By comparing native speaker and Brazilian learner data, one can infer that the Brazilian learners of this study have not yet incorporated the discourse and pragmatic functions of "*actually*" into their production. Only 3 learners produced "*actually*": 3 occurences in utterance initial position, 1 occurrence in medial position. No occurrence of "*actually*" final was observed in the Brazilian data, the most common position as a discourse marker in native speaker production.

Taking into account all results, it is possible to conclude that the subjects of this study would benefit from some pedagogical assistance for acquiring discourse markers in general, based on the limited range of discourse markers observed in their spoken production and on the underuse of the pragmatic and discourse functions of "*well*" and "*actually*". The Brazilian learners of this study produced the discourse marker "*well*" mostly as a hesitating device and with a considerably lower frequency than that of native speakers. Few occurrences of attitudinal "*well*" were observed. As for "*actually*", only 4 occurrences were observed in the Brazilian corpus and none of which had been placed in utterance final, the most common position of "*actually*" as a discourse marker in native speaker production.

As for pedagogical recommendations, Thornbury and Slade (2006, p. 296) suggest that an approach to the teaching of conversation should include three elements, namely exposure, instruction and practice, not necessarily in this order. This approach is aligned with Schmidt's (1993) theory that accounts for pragmatic learning (previously presented in section 2.1). His main claims are as follows (Schmidt, 1993, pp. 35-36): attention to input is a necessary condition for any learning itself;for the learning of pragmatics in a second language, attention to linguistic forms, functional meanings, and the relevant contextual features is required; learners experience their learning; attention is subjectively experienced as noticing; the attention threshold for noticing is the same as the threshold for learning; while incidental and implicit learning are both possible, consciously paying attention to the relevant features of

input and attemping to analyse their significance in terms of deeper generalisations are both highly facilitative; explicit teacher-provided information about the pragmatics of a second language can also play a role in learning, on condition that it is accurate and not based solely on fallible native speaker intuitions.

Thornbury and Slade (2006, p. 296) state that in an exposure-instruction- practice cycle, learners may listen to conversation extracts, study specific features highlighted in transcripts and then attempt to use them in their own conversations. In an instruction-exposure-practice cycle, learners are given prior explicit instruction in a feature of conversation. In an exposure-practice-instruction cycle, students receive feedback (instruction) after they have listened to conversation extracts and tried to implement specific features. A practice-exposure-instruction cycle encompasses attemping a conversational task, observing more proficient speakers performing the same task and receiving instruction, before repeating the task. The authors conclude that the sequence of events is less important than the fact that all the three elements are included in an instructional programme.

Considering discourse markers in particular, teachers can previously select authentic conversation extracts that include a range of discourse marking adverbs and that are at a level of listening proficiency their learners can cope with. In class, learners listen to selected extracts and initially work with tasks aimed at comprehension (exposure). Once overall meaning has been established, learners receive transcripts of extracts so as to study the discourse functions of specific adverbs. To make analyses easier and to save class time, it is also a good idea to create a matching exercise that lists the functions of specific discourse markers (instruction). Once functions have been studied, learners can engage in similar conversation activities that resemble the situation type portrayed in extracts and that give learners the opportunity to use the proposed discourse markers (practice).

Reordering elements, one can also consider a practice, exposure and instruction (and practice) cycle, also known as a test-teach-test approach. Focusing on long turns in particular, learners can be given a task in which they are expected to talk about a topic for about 1 to 2 minutes (practice). Alternatively, learners can also use their mobiles to record their production for future reference. Students then listen to a similar extract in which a more proficient learner or a native speaker performs the same task and uses a range of discourse marking adverbs

(exposure). After that, students receive extracts, discuss major differences vis-à-vis the use of discourse markers and study their functions (instruction). A final stage is to ask learners to perform the task again, this time using the discourse markers analysed (practice). They may also record their production again for future reference and comparison.

Before concluding this section, let us relate findings and recommendations to the intermediate plateau phenomenon. Instruction on the discourse and pragmatic functions of discourse markers seems to be a relevant teaching goal to help learners move beyond the intermediate plateau.As previously mentioned, Science without Borders applicants need to obtain a successful performance at CEFR B2 or C1, depending on the demands of the target university or on the examination taken per se. Candidates taking Cambridge English First (CEFR B2) and Advanced (CEFR C1) are assessed in four and five analytical criteria, respectively. Considering the discourse management criterion (as previoulsy detailed in section 3.3.2), a range of discourse makers is expected from candidates at CEFR B2 and C1. Thus, candidates' command of discourse markers plays a role in the assessment of their speaking ability.

Lastly, the qualitative analyses of this chapter have provided me with some insight into assessment. Gilbert and Staub (2014) report on their investigation into speaking examiners' confidence when making assessment judgements in relation to Cambridge English Advanced (CEFR C1). The participants of their study were all experienced speaking examiners and reliable raters in Switzerland. One proposed task was to ask respondents to rate their confidence (from very confident to not confident) in applying each of the analytical scales across the full range of speaking tests (Gilbert and Staub, 2014, p. 53). While interactive communication (described in section 3.3.2) was the criterion they felt most confident with (87.2%), discourse management had the lowest percentage (72.5%). Code components were as follows: grammar (85.2%), vocabulary (83.2%) and pronunciation (81.1%). Examiners commented on their reasons for less confidence with the discourse management scales (Gilbert and Staub, 2014, p. 54):

- "*I would like to be clearer about how to interpret the IC and DM scales across the CEFR levels*".
- *With DM I often have a problem with uses a range of cohesive devices. Very often candidates don't use cohesive devices but still have good DM"*.

- "*The descriptors for DM are useful in identifying a level higher than FCE (i.e. CEFR C1 and C2). The idea of Discourse Markers is introduced here and that gives me something to hold on to*".

In my view, the wide range of functions that discourse markers (and pragmatic markers in general) perform may also pose difficulties in applying the discourse management criterion. While some discourse markers have clearly defined discourse functions, i.e. mediating segments of discourse and marking speakers' stance towards the message, other discourse or pragmatic markers may be used as strengthening or mitigating devices, affecting the illocutionary force of utterances. In other words, conveying different shades of pragmatic meaning. In addition, the analytical criterion "interactive communication" presupposes that candidates initiate and respond appropriately and maintain the interaction. Thus, there may be some overlap since a number of discourse markers are used to initiate and to close conversations. Discourse markers may also be employed as listener response tokens as a means to give feedback to speakers on how the message is being received and to encourage them to continue their ongoing turns. The next two chapters of this thesis address some overlapping functions of pragmatic markers in the light of the proposed model of conversational competence.

To sum up, in this section we have summarised corpus-based findings vis-à-vis the relative frequencies and the discourse functions of the adverbs "*well*", "*really*" and "*actually* in the Brazilian learner corpus and in native speaker corpora. We have discussed these findings in the light of the specific research sub-question addressed in this chapter. We have proposed some pedagogical recommendations and related them and results to the intermediate plateau phenomenon. We have also highlighted the wide range of functions that discourse markers and pragmatic markers perform.

SUMMARY

This chapter has investigated the most common discourse marking adverbs used to mediate segments of discourse in conversations by Brazilian learners at CEFR B1. While most discourse markers belong to the general class of adverbs, it is their function as organisers of larger stretches of discourse that qualifies them as such (Carter & McCarthy, 2006). Stance

adverbs mark speakers and writers' personal stance, i.e. the expression of personal feelings and assessements in three major domains (Conrad & Biber, 1999). Epistemic stance comments on the certainty (or doubt), reliability, or limitations of a proposition; attitudinal stance conveys the speakers' attitudes, feelings, or value judgements; style stance describes the manner in which the information is being presented.

AntConc's word list tool was used to view the rank positions and the raw frequencies of all the words in the Brazilian learner corpus. Results indicated that the adverbs "*maybe*", "*really*", "*just*", "*well*", "*probably*" and "*actually*" seem to be the most common discourse marking adverbs used by the subjects of this study. Manner adverbs, frequency adverbs and adverbs with fewer occurrences than 4 were disregarded. The adverbs "*really*", *"well*" and "*actually*" were selected for investigation in this chapter since they are commonly used to mediate segments of discourse with regards to the different parts of discourse and also to the relations between speakers and parts of discourse. "*Well*" is a common and frequent single word adverb used to facilitate opening and closings of conversations (Carter & McCarthy, 2006) and "*really*" and "*actually*" are highly frequent stance adverbials used to mark speakers' attitude or stance to the message (Conrad & Biber, 1999).

CQP's frequency list tool was used to display the frequency of word forms in the benchmark corpora. The data attested high relative frequencies of "*well*", "*really*" and "*actually*", indicating ever higher rank positions and relative frequencies in the specialised sub-corpus of face-to-face conversations, and thus, strengthening their association with face-to-face conversations. It was also highlighted that "*well*" was more frequent than "*really*" in British English native speaker production while the Brazilian learners of this study produced "*really*" more frequently than "*well*".

Chapter sub-sections investigated the pragmatic functions and frequencies of the items under investigation in native speaker and learner corpora. Aijmer (2011) categorises "*well*" into speech management and attitudinal functions. There were 9 occurrences of "*well*" in the Brazilian learner corpus. Subjects used "*well*" mostly for speech management functions, namely choice and change-related functions. Attitudinal functions were less frequent. The subjects of this study used "*well*" more frequently in the choice-related function than in the change- related function. The Brazilian data analysis revealed an uneven distribution of "*well*"

among interviews and subjects. Only 3 participants included "*well*" in their oral production. This possibly indicates lack of familiarity with the discourse functions of "*well*" by the Brazilian learners of this study, and thus, it may signal underuse. In addition, "*well*" was mostly used for speech management functions and often preceded by filled and unfilled pauses in initial position or in medial position. These results corroborate the assumption that "*well*" was mostly used as a floor holding device, similarly to Aijmer's results (2011).

Paradis (2003) reveals three functions of "*really*" based on semantic analyses in COLT: truth attesting, emphasising and degree reinforcing. Subjects were able to produce "*really*" so as to perform all the three pragmatic functions proposed by Paradis (2003), including the de-emphasiser one, though overall relative frequencies differed. The Brazilian learners displayed a higher percentage of emphasising "*really*" and a lower one of truth attesting if compared to those in native speaker data. This difference may signal that the subjects of this study tend to produce "*really*" more frequently within utterances, either to emphasise scalar properties of adjectives or to reinforce the subjectivity of situations when accompanying attitudinal verbs. As for the distribution of "*really*" in the Brazilian learner corpus, 8 out of 10 interviews included "*really*". The comparison between Brazilian and Spanish learners showed that both nationality types produced "*really*" inserted within utterances. The Spanish learners produced "*really*" with a higher frequency, even higher than that of native speakers.

Aijmer (2002) draws a distinction between "*actually*" as a regular adverb and as a discourse particle based on utterance position. When "*actually*" is placed medially, it hedges the proposition or a single element. When it is placed initially or finally, it is a discourse participle with textual or interpersonal function. Aijmer's (2002) investigations in the LLC indicate that the most typical use of the discourse particle is in final position, being three times more frequent than in initial position. The Brazilian learner data showed that "*actually*" was mostly used as a discourse particle, since 3 out of 4 occurrences took place in initial position. All 4 occurrences were followed by "*I*" and verbs expressing preferences. In the examples of the discourse particle, "*actually*" was used to introduce one's opinion or preference conveying different attitude to the preceding discourse.The single occurrence of "*actually*" in medial position displayed its use as a mitigating device. It was suggested that the absence of "*actually*" in utterance final may signal lack of familiarity with its functions in this

position. Spanish learner results (LINDSEI) indicated that the Spanish produced "*actually*" with a higher frequency than that of the Brazilian learners.

The final part of this chapter discussed "*well*", "*really*" and "*actually*" findings in the light of the specific sub-research question addressed in chapter 4. All occurrences of "*well*" in the Brazilian data portrayed its use as a discourse marker, mediating segments of discourse. The subjects of this study used "*really*" more frequently as a pragmatic marker modifying speaker meaning, considering the degree reinforcer, emphasiser and de-emphasiser functions. Brazilian learner data indicated only 4 occurrences of "*actually*", 3 of which as a discourse marker in initial position. It was concluded that the subjects of this study would benefit from some pedagogical assistance for acquiring discourse markers in general, based on the limited range of discourse markers observed in their spoken production and on the underuse of the pragmatic and discourse functions of "*well*" and "*actually*". Thus, pedagogical recommendations based on a three-part cycle (Thornbury & Slade, 2006) were proposed, aligned with Schmidt's view (1993) on pragmatic learning. It was also suggested that instruction on the discourse and pragmatic functions of discourse markers offers learners an opportunity to move beyond the intermediate plateau.

Let us now turn our attention to the second facet of conversational competence, namely the negotiation of illocutionary meaning, by focusing on adverbial hedges used to mitigate representative speech acts.

5. MITIGATING REPRESENTATIVE SPEECH ACTS: EXPLICIT AND IMPLICT ADVERBIAL HEDGES

According to Caffi (2013, p. 277), "speakers mitigate out of certainty, caution and consideration". Mitigation is pervasive in speech since speakers commit themselves to only what is strictly necessary. Mitigation is not only a matter of politeness as it captures rationally grounded behaviour mainly aimed at avoiding unnecessary risks, responsibilities and conflicts. Therefore, mitigation is a bridging category between different paradigms, being a cognitive, a relational and an emotive category (Caffi, 2013).

In chapter 1.2.3, it was stated that term "hedge" can be used to refer to both softening and strengthening devices (Brown & Levinson, 1987) or to softening functions only (Prince et al., 1982). For the purposes of this chapter, I adopt Nikula's (1996) characterisation of pragmatic force modifiers, in which "hedges" are viewed as a sub-type of pragmatic modifiers, used for mitigating and softening purposes. Furthermore, the most recurrent class of speech acts (Searle, 1976) performed by the learners of this study were "representative acts" such as asserting, speculating and concluding. These acts commit the speaker to the truth of the expressed proposition (Levinson, 1983). Therefore, I chose to investigate forms of hedges used to mitigate the degree of assertiveness or to lower the degree of certainty in contexts of assertions, deductions and conclusions.

This chapter addresses the following research question, focusing specifically on its sub-question:

RQ2: *How do Brazilian learners at CEFR B1 mitigate representative speech acts?*
RSQ2: *What are the most common explicit and implicit adverbial hedges used to mitigate representative speech acts by Brazilian learners at CEFR B1? How do their frequency of use and pragmatic functions compare to those of native speakers?*

In the first section of this chapter, I briefly comment on Nikula's (1996) distinction between explicit and implicit pragmatic force modifiers and then present corpus-based evidence (O'Keeffe et al., 2007) of forms traditionally associated with hedging. Next, I introduce the possible candidates for hedges in the Brazilian learner corpus, discuss their main functions

and justify choices for adverbial selections. Lastly, I present benchmark native speaker data and the most common functions of the selected items for investigation, namely the pragmatic markers "*probably*", "*maybe*" and "*just*".

In the sub-sections of this chapter, I investigate the pragmatic functions of the selected adverbial hedges in the Brazilian learner corpus and compare frequencies. I introduce each sub-section with a categorisation of the pragmatic functions of the item under investigation (Chafe, 1986, Palmer, 2001, Biber et al., 1999, Wierzbicka, 2006, Aijmer, 2002) and related corpus-based findings. Next, I present the Brazilian learner corpora data (position, frequency and functions), qualitative analyses and comparisons with the native speaker benchmark corpora and the Spanish component of LINDSEI. Some elements display low raw frequencies in the Brazilian learner corpus. As a consequence, my comments can only be indicative of general patterns.

In the final part of this chapter, I summarise the main points from the previous sections, discuss findings in the light of the proposed research sub-question and make some pedagogical recommendations.

5.1. HEDGING FORMS

As previously discussed in section 1.2.3, pragmatic force modifiers refer to modifying devices that speakers use either to soften or to strengthen the force of their messages (Nikula, 1996). Pragmatic force relates to context-dependent functions that an utterance can have, apart from its semantic meaning. In this section, I shall be using the term "pragmatic force modifiers" to specifically refer to devices that soften or mitigate the pragmatic force of utterances, namely "hedges". Nikula (1996, pp. 50-55) draws a distinction between explicit and implicit pragmatic force modifiers based on the semantic-pragmatic divide. Explicit modifiers include attitudinal adverbs, parenthetical constructions or adverbs of degree. They indicate quite explicitly the degree to which speakers are committed to the truth or preciseness of their messages. Implicit modifiers, on the other hand, have a tendency to remain ambiguous or fuzzy even in the context, often making various interpretations possible.

Corpus-based studies provide evidence of different forms used for mitigating purposes. O'Keeffe, McCarthy and Carter (2007, p. 175) categorise the most common forms of hedges (in the sense of pragmatic markers) found in the Limerick Corpus of Irish English (LCIE):

a) Modal verbs and verbs with modal meaning: "***I guess*** *the bus service isn't too good, is it?"*

b) Nouns: "*there's a* ***possibility****, the* ***thing*** *is, etc*"

c) Adverbs: degree – "***quite***", "***really***", "***relatively***", "***necessarily***"; restrictive- "***just***", "***only***"; stance- "***of course***", "***actually***", "***kind of***", "***sort of***", "***really***", "***maybe***"

d) Syntactic choices: choice of question form – "***And would you have thought*** *you were very close to him?*"; double negative – "***it's not*** *that* ***I'm not*** *afraid*"; evaluation relative clause insertion- "*You got them to do this cross-group reporting,* ***which was a good idea,*** *but the time was the problem*"

e) Features of "onlineness" adjustments (false starts and repetitions): "*And* ***will you would you*** *like to go sort of on a sun and sea holiday with him this year*?"

O'Keeffe, McCarthy and Carter (2007, p. 175) also report on the 10 most frequent hedging items in the Limerick Corpus of Irish English: "*just*", "*really*", "*actually*", "*probably*", "*I think*", "*a bit*", "*kind of*", "*sort of*", "*you know*" and "*I suppose*". Applying Nikula's (1996) notion of an explicit-implicit continuum, more explicit forms possibly include "*I think*", "*probably*" and " *I suppose*" whereas forms that are likely to trigger different interpretations, and thus may be placed at the implicit end of the continuum, are "*just*", "*really*" and "*you know*".

Let us now examine the possible candidates for hedging forms in the Brazilian learner corpus, considering mitigating or softening purposes. I used AntConc's word list tool to view the rank positions and raw frequencies of all the words in the Brazilian learner corpus. As the corpus is not annotated, I made a manual list of all the lexical items that fit into the form categorisation of hedges proposed by O'Keeffe, McCarthy and Carter (2007) and/or are traditionally associated with mitigation. Items with fewer than 4 occurrences were disregarded. The adverbs "*well*", "*really*" and "*actually*" were not considered since they had been analysed as discourse marking adverbs used to mediate segments of discourse in the previous chapter. I then conducted individual searches of the selected items and sorted results so that sentence patterns and collocations became more salient. Following this, I disregarded the occurrences

of selected items that had been produced by the interlocutor and compiled results. Table 11 includes the raw frequencies (actual number of occurrences) and the relative frequencies (per 10,000 words) of the lexical items that fit the form characterisation of hedges as proposed by O'Keeffe, McCarthy and Carter (2007, p. 175) and/or are traditionally associated with mitigation.

Table 11: Most frequent lexical items fitting the form characterisation of hedges in the Brazilian learner corpus

Lexical item	Raw frequencies	Relative frequency
think	245	102.9282
can	134	56.2954
things/thing	61/ 21= 82	34.4494
maybe	71	29.8281
about	63	26.4672
would	53	22.2661
will	46	19.3252
kind	42	17.6448
just	17	7.1419
seems/seem	11/6 = 17	7.1419
probably	13	5.4614
feel	10	4.2011
could	9	3.7810
(in) my opinion	9	3.7810
guess	8	3.3609
bit	6	2.5206
should	5	2.1005
types/type	4	1.6804

The above results indicate that the Brazilian learners of this study produced a range of word forms traditionally associated with mitigation. Forms associated with hedging included attitudinal verbs, modal verbs, nouns and epistemic and restrictive adverbs. However, not all selected forms were necessarily used for mitigating purposes. In order to examine the most common uses of the lexical items, I analysed concordance lines of individual items. When there were too many samples, I relied on the sorting tool so as to make patterns more salient.

Table 11 shows that the attitudinal verb "*think*" was the form associated with mitigation most frequently produced by the Brazilian learners of this study. 215 occurences (90.3247 relative

frequency per 10,000 words) out of 245 included the collocation "*I*" + "*think*". This formula is widely used by Brazilian learners in general as a means to introduce their opinion rather than as a hedging device. Another typical form used to introduce learners´ opinion is the fixed expression "*In my opinion*", also present in the Brazilian learner corpus, but with a lower frequency. In addition, occurrences of the verb "*guess*" were also preceded by "*I*", displaying its use for introducing one's opinion, but conveying a much lower degree of certainty with regards to the truth value of the subsequent proposition.

As for the second most frequent form, the modal verb "*can*" was found mostly in contexts of suggestions, expressing possibility and ability. Less assertive forms within suggestions included the other modal verbs "*could*" and "*should*", although such distal forms were used with a much lower frequency. "*You can*" was the most common collocation with 71 occurrences (29.8281 relative frequency per 10,000 words) expressing mostly possibility. Only 14 occurrences of "*I can*" were found, out of which 11 expressed ability (mostly followed by sensory verbs as a marker of sensory evidence) and 3 ocurrences conveyed possibility.

The third most frequent form was the lemma "*things/ thing*". All occurrences of "*things*" displayed its use for expressing vagueness within categories. I use the term "vagueness" in a similar way to Chafe's notion of hedges (1986, p. 270) as devices to make categories less precise. The most common collocating items with "*things*" were as follows: "***the*** *things*" (11 occurrences), "***a lot of*** *things*" (4 occurrences) and "***some*** *things*" (3 occurrences). Other forms for expressing vagueness included "*kind of*", "*a bit*", "and "*types/type*", but with lower frequencies. As for the singular form "*thing*", it also expressed vagueness.

Now that we have presented the most common uses of the three most frequent forms displayed in table 11, let us illustrate them with authentic data from the Brazilian learner corpus.

(1)

<S16> I like to live in Porto Alegre because I like the people too. **I think** they are really smart and receptive uhm, uhm principally in my age and I like, **I think** the city is not too beautiful but have many **things** to do, cultural **things** and parks and **things** like that, many uh like restaurants, pubs and bars and **things**, uh places to go out.

<S1>Do you prefer to spend time on your own or with other people?
<S16> **I think** with other people, with my friends.
(FCE P1- B1 BR-08)

In conversation (1), one can notice the recurrent use of "*I think*" as a fixed formula for expressing one's opinion in utterance and clause initial positions. The last example of "*I think*" is followed by a prepositional phrase rather than the projected clause type (which was omitted). Subject 16 also produces 4 occurrences of "*things*" within the same long turn, expressing imprecision of categories. While "*many things to do*" and "*things like that*" refer to a broarder category of activities that can be done in Porto Alegre, the noun phrase "*cultural things*" indicates a vague category of cultural activities. The last example of "*things*" seems to refer to a vague category of places to go out. All these occurrences within the same turn seem to help the speaker to sustain his turn.

(2)
<S1> Thank you. <SE> pause </SE> Would you like to spend time in a café like this?
<S17> Yes! Of course!
<S16> Yes, it's really nice you, when, where you **can** do a lot of **things**, you **can** read, you **can** uh have a conversation, you **can** meet uh persons and I really like to +
<S17> <SE> overlapping </SE> You **can** watch TV.
<S16> + Yeah, watch, especially in a café.
(FCE P4- B1 BR-08)

In conversation (2), there are 5 occurrences of "*can*" expressing possibility. One occurrence of "*things*" is used as a vague category of possible activities to be done in a café. It also introduces the idea that there a lot of activities to be done in a café. Such activities are then specified in the subsequent clauses of the same speaker, all preceded by the modal verbal "*can*".

As for adverbial forms in particular, "*maybe*", "*probably*" and "*just*" were found mostly within contexts of representative speech acts such as asserting, making deductions and concluding. For the purposes of this thesis, I chose them for the empirical investigations of hedges for the following reasons. First of all, "*maybe*", "*probably*" and "*just*" are adverbs. As

such, they can be placed in different utterance positions, displaying more versatility than attitudinal verbs and modal verbs, which are necessarily attached to preceding subjects. Thus, more varied uses may be observed. Secondly, the previous analyses of the most common collocations with "*things/thing*" revealed their uses attached to vague categories within noun phrases. Thus, such uses are more limited in scope vis-à-vis the mitigation of representative speech acts. Thirdly, "*maybe*" and "*probably*" are epistemic adverbs traditionally associated with degrees of certainty of propositions while "*just*" may be employed as a politeness strategy of speech acts in general. Therefore, they are likely to be employed for mitigating the degree of assertiveness or to lower the degree of certainty in contexts of assertions, deductions and conclusions.

In addition, while "*maybe*" and "*probably*" seem to be more explicit forms of hedges, the pragmatic marker "*just*" appears to be more ambiguous, lending itself to different interpretations. Thus, these adverbs seem to offer a richer opportunity for pragmatic analyses. Lastly, the Brazilian corpus does not display a single occurrence of the modal verb "*may*", which may suggest a tendency to use adverbial equivalents to convey this degree of certainty instead of the modal verb form. There were 11 occurrences of the distal form "*might*" but all produced within the interviewer's frame. The investigation of "*maybe*" and "*probably*" may provide interesting data related to the absence of both modal verbs in the Brazilian spoken production.

Before we move on to focus on the frequencies and on the functions of "*maybe*", "*probably*" and "*just*", let us summarise the main uses of the other lexical items in the Brazilian learner corpus not yet referred to.

a) The preposition "*about*": mostly used as a dependent preposition, collocating with the verb to be "*…is about*" and verbs in general "*learn/talk/read/take notes about*";

b) The modal verb "*would*": largely employed within answering moves, echoing part of the interlocutor's question, which included the modal verb. Collocating items included "*I would like/love/prefer*" (expressing positive feelings) and "*it would be a good idea*" and "*It would be cool/good/ nice*" (in contexts of suggestions);

c) The modal verb "*will*": mostly used to convey futurity. Collocating items included "*will go/return/spend/work*" (future plans), "*will be/ attract/make/ not be/not attract*" (for future predictions) and "*I will talk about*" (for spontaneous decisions);

d) The verb "*feel*": used for expressing both positive and negative feelings "*I feel happy/ good/ relaxed/ attracted/lonely*".

Returning our attention to "*maybe*", "*probably*" and "*just*", let us investigate their frequency in the native speaker benchmark corpora. Table 12 summarises the raw frequencies and the relative frequencies (per 10,000 words) of "*maybe*", "*probably*", "*just*" in the BNC Sampler spoken sub-corpus, in the face-to-face conversations from the 1990s and in the Brazilian learner corpus (occurrences produced by the interviewer have been disregarded).

Table 12: "*Maybe*", "*probably*" and "*just*" frequencies in 3 corpora

	BNC		**F-2-F**		**Brazilian corpus**	
	Raw frequencies	Relative frequency	Raw frequencies	Relative frequency	Raw frequencies	Relative frequency
Maybe	309	2.7103	98	3.6114	71	29.8281
Probably	524	4.5962	218	8.0336	13	5.4614
Just	3,884	34.0685	1,168	43.0424	17	7.1419

The above data display an extremely high frequency of "*just*" in both benchmark corpora, which possibly corroborates the assumption that "*just*" is an implicit pragmatic force modifier that lends itself to a number of uses and interpretations. As for the epistemic adverbs, "*probably*" occurs twice as often as "*maybe*" in the specialised spoken sub-corpus. It is also more frequent than "*maybe*" in the BNC Sampler spoken sub-corpus. By comparing native speaker results with the Brazilian data, one can infer that "*maybe*" was possibly overused by the subjects of this study while the occurrences of "*just*" in the Brazilian corpus may indicate some possible underuse. Another salient feature is that the Brazilian learners of this study used "*maybe*" more frequently than "*probably*", differing from British native speaker linguistic behaviour. But before drawing premature conclusions, let us present what the literature says about the most common functions of "*maybe*", "*probably*" and "*just*".

Biber et al. (1999, pp. 867-868) state that all the most common stance adverbials mark epistemic stance. They indicate that three adverbials are especially common across all registers: "*of course*", "*perhaps*" and *probably*". The common use of "*perhaps*" and

"*probably*" in all registers highlights the function of stance adverbials as a means to show the doubt or certainty of the proposition. Both adverbs occur commonly in conversations, yet "*probably*" occurs twice as often as "*perhaps*" (Biber et al., 1999, p. 868):

(3) "*We can **probably** leave packing until next week*".

(4) "***Perhaps** I'll have two, cos I'm a pig*".

Biber et al. (1999: 868) stress a difference between British and American conversation with regards to epistemic adverbs. While the adverb "*maybe*" is used in both dialects, it displays a much higher frequency in American English, similarly to the results in the Brazilian learner corpus. Biber et al. (1999) indicate that speakers use "*maybe*" to mark uncertainty in many situations where "*probably*" and "*perhaps*" could also be used, including giving possible explanations for events, interpreting the outcome of events, and considering future actions. The below utterances are all examples of American conversation (Biber et al., 1999, p. 868):

(5) "***Maybe** she was in California*".

(6) "*Well, but **maybe** it's good that he hears this right now instead of at, in his twenties*".

(7) "***Maybe** I'll put half of 'em out in the beginning and half of 'em in the middle*".

As for restrictive adverbials, Biber et al. (1999, p. 780) state that such adverbs emphasise that the proposition is true in a way which expressly excludes some other possibilities. Restrictive adverbs cannot be moved without affecting the meaning of the clause. Furthermore, in some cases, adverbials combine qualities of both the restrictive and extent/degree categories. The adverb "*just*" is the most common single restrictive adverb in conversations (both American and British varieties). It often appears to have qualities of both restricting the action and lessening the intensity. The following examples illustrate the different functions and the versatility of "*just*" (Biber et al., 1999, pp. 798-799)

(8) "*I **just** can't believe it".*

(9) "*Now, now **just** sit down!*"

(10) "*Oh I **just** wondered- **just** wondered who it could be*".

Example (8) shows the primary role of "*just*" as a restrictive adverb focusing on the part of the clause for which the truth value of the proposition is most important. In example (9), "*just*" is used within an imperative utterance conveying a strong sense of "Don't argue: simply do as I say". In example (10), "*just*" has the effect of softening what is being said.

Biber et al. (1999, p. 799) explain that this function particularly occurs in contexts in which speakers are justifying their or other people's actions.

To sum up, in this section we have presented forms traditionally associated with mitigation, introduced the possible candidates for hedges in the Brazilian learner corpus and justified choices based on corpus evidence. O'Keeffe, McCarthy and Carter (2007, p. 175) categorise the most common forms of hedges found in LCIE: modal verbs and verbs with modal meaning, nouns, degree and stance adverbs, syntactic choices, double negative, evaluation relative clause insertion and features of "onlineness" adjustments (false starts and repetitions). The investigation of the possible candidates for hedging forms in the Brazilian corpus indicated that the subjects of this study produced attitudinal verbs, modal verbs, nouns and epistemic and restrictive adverbs. "*Maybe*", "*probably*" and "*just*" were found mostly within contexts of representative speech acts such as asserting, making deductions and concluding. The relative frequencies of the three adverbs were compared to those in native speaker data. Following from this, the epistemic and restrictive adverbs were selected for empirical investigations, to be detailed in subsequent sub-sections.

5.1.1. "*Probably*" and "*maybe*"

This section briefly introduces views on evidentiality (Chafe, 1986), evidential and epistemic modals (Palmer, 2001), epistemic stance markers (Biber et al., 1999) and on the semantic meanings of "*probably*" (Wierzbicka, 2006). It then presents Brazilian learner corpora data (position, frequency and functions), qualitative analyses and comparisons with other corpora.

As previously mentioned in chapter 1.2.3., in a broad sense, evidentiality refers to linguistic expressions of attitude to knowledge. Chafe (1986) indicates that people are usually aware that some portions of their knowledge are truer or more accurate than others and that not all knowledge is equally reliable. Thus, speakers qualify such knowledge with expressions that indicate their assessment of its degree of reliability, the likelihood of its being a fact. Chafe (1986, pp. 264-265) points out that conversational English uses adverbs and modal verbs for this purpose:

(1) "*We kept thinking **maybe** they'd be stationed at the Presidio*".

(2) "*But I'm **probably** not going to do it anymore*".

(3) "*He's quite nice, and **certainly** very cheery*".

(4) "*The one thing that **might** shed light on it was something about requests*".

(5) "*You know I **may** not really be putting in full time even this quarter*".

Current views on modality draw a distinction between evidential and epistemic modals. Palmer (2001) explains that epistemic and evidential systems are two main types of propositional modality. "With epistemic modality speakers express their judgements about the factual status of propositions whereas with evidential modality they indicate the evidence they have for its factual status" (Palmer, 2001, p. 8). Within epistemic modality, there are three types of judgements that are common in languages:

a) Speculative: it expresses uncertainty;
b) Deductive: it indicates inference from an observable evidence;
c) Assumptive: it indicates inference from what is generally known.

From a corpus-based perspective, Biber et al. (1999, p. 972) state that "epistemic stance markers are used to present speaker comments on the status of the information in a proposition". Such markers can indicate certainty or doubt, actuality, precision, or limitation. They can also indicate the source of knowledge or the perspective from which the information is given. Biber et al. (1999) clarify that many individual stance markers convey only a single stance meaning, for instance, the adverb "*probably*" provides an assessment of certainty or doubt. Conversely, some stance markers may not necessarily convey just one single meaning, for example, the verb "*think*' controlling a that-complement clause can mark the degree of certainty and indicate the source of knowledge at the same time.

Biber et al. (1999, pp. 972-973) detail the forms epistemic markers can take as a means to express certainty or doubt, actuality, precision or limitation:

a) Adverbial: single adverbs ("*definitely*", "*possibly*", "*maybe*", "*probably*", etc.), prepositional phrases ("*in fact*", "*without doubt*", etc.), hedge ("*sort of*" and "*like*") and comment clause ("*I think*");
b) Verb/adjective/noun + complement clause: "*I **know** I can get off the bus*"; "*In fact I'm not **sure** that they did very much at all*"; "*There was a **suggestion** that the bidder may be a financial buyer*".

c) Stance noun + prepositional phrase: "*But there is a **real possibility** of a split within the Lithuanian party*".
d) Modal verb in extrinsic sense: "*I think you **might** be wrong*"; "*Legumes **may** have smaller conversion efficiency than cereals*".

Wierzbicka (2006, p. 261) states that "*probably*" is the most important of all English epistemic adverbs, considering its high frequency and spread across a wide range of registers and types of discourse. The differences in frequency between "*probably*" and other epistemic adverbs are quite spectacular. The frequency of "*probably*" in spoken English is 28 tokens per 50,000 in the Cobuild UK spoken corpus while it is 44 tokens per 50,000 words in the Lund corpus. With regards to the semantic meaning of the adverb "*probably*", Wierzbicka (2006) argues that it should not be uncritically equated with that of the adjective "*probable*" as they cannot always be used interchangeably. Firstly, "*probably*" can be used with personal opinions and moral judgements whereas "*probable*" cannot (Wierzbicka, 2006, p. 266):

a) "*It's **probably** not worth it*" versus "It is probable that it is not worth it";
b) "*I **probably** should do it*" versus "It is probable that I should do it";
c) "*You **probably** ought to try harder*" versus " It is probable that you ought to try harder".

Secondly, "*probable*" can be modified by the adverb "*highly*" whereas "*probably*" cannot, for instance "*highly probable*" versus "highly probably". Thirdly, collocations with "*highly probable*" suggest that something is somehow measurably quantifiable as one can ask questions like "*how probable something is*" or "*what is the probability that...*". Wierzbicka (2006) clarifies that these questions do not apply to "*probably*" as it is not an estimate but a hedge. Lastly, "*probably*" has a reduced variant in spoken English "*prob'ly*" and its frequency is incomparably higher than that of "*probable*".

Based on the above differences, Wierzbicka (2006) argues that "*probably*" is less scientifical and intellectual than "*probable*" and that "it is well and truly entrenched in English colloquial speech". "*Probably*" is epistemically cautious and implies a certain amount of conscious reflection, e.g. perhaps some mental weighting of the evidence. However, it can be relatively causal as it does not imply that one can prove the validity of what one is saying or

demonstrate how one has arrived at one's judgement. Thus, Wierzbicka (2006, p. 266) decomposes the meanings of "*probably*" as follows:

(6) "*Probably he did it himself*" =

a) I want to say how I think when I think about it;
b) I say: he did it himself;
c) I don't say I know;
d) I think it can be like this;
e) I think like this because I have thought about it for some time;
f) I think that other people can think the same.

Wierzbicka (2006) explains that like other epistemic adverbs, "*probably*" indicates that one thinks something and also that one abstains from claiming knowledge. The assertiveness of "I say : he did it himself" is counterbalanced by the cautious assumption "I think it can be like this". At the same time, the speaker indicates that this reasoning is based on reflection: "I think like this because I have thought about it for some time". Such confidence is also reflected in "I say: he did it himself". Wierzbicka (2006, p. 267) concludes that "it is this component of assertion (however qualified by the other components) that gives the adverb "*probably"* a certain casual confidence, which is absent from the more scientific-sounding adjective "*probable*".

Let us now examine the distribution of the epistemic adverbs "*probably*" and "*maybe*" in the Brazilian learner corpus. Table 13 displays the raw and relative frequencies (per 10,000 words) of "*probably*" and "*maybe*" in utterance initial, medial and final positions. 1 occurrence of "*maybe*" has been disregarded (produced by the interviewer).

Table 13: Position distribution of "*probably*" and "*maybe*" in the Brazilian learner corpus

Position	Probably	Relative frequency	Maybe	Relative frequency
Utterance initial	X	X	9	3.7810
Medial position	12	5.0413	58	24.3666
Utterance final	1	0.4201	4	1.6804
Total occurrences	**13**	**5.4614**	**71**	**29.8281**

The above results indicate that the Brazilian learners of this study produced both epistemic adverbs "*probably*" and "*maybe*" mostly in medial position, despite differences in individual relative frequencies. In other words, the epistemic adverbs were placed within utterances.The second most frequent position was "*maybe*" utterance initial as there were no occurrences of "*probably*" in this position. Both adverbs were also found in utterance final position with lower frequencies. Overall, the relative frequency of "*maybe*" is considerably higher than that of "*probably*", indicating a possible favouritism towards the epistemic adverb "*maybe*" by the Brazilian learners of this study. Different reasons may account for this tendency. Firstly, as stated by Biber et al. (1999), "*maybe*" is much more frequent in conversations in American English than in British English. American speakers tend to use "*maybe*" to mark uncertainty in many situations where "*probably*" and "*perhaps*" could also be used. Following from this, Brazilian learners in general have much more exposure to the American variety due to various reasons: Brazil is geographically closer to the USA, which may facilitate more frequent visits to the USA; most Brazilian schools (both regular schools and language institutions) adopt American English textbooks; a regular and more frequent access to American movies and sitcoms. Secondly, as the Brazilian learners of this study did not produce any occurrences of "*may*" and "*might*" for expressing degrees of certainty , this may indicate that the subjects of this study resorted to the epistemic adverb "*maybe*" (and possibly *"probably*" with a much lower frequency) to perform this function.

Let us now describe some contexts of usage of both epistemic adverbs in utterance initial, medial and final positions. 6 occurrences of "*maybe*" in utterance initial position were found in contexts of discussions, where learners had been considering the pros and cons of suggestions for a café. Overall, "*maybe*" was employed as a mitigating device to mark the degree of uncertainty vis-à-vis the effectiveness of the proposed suggestions. 2 examples of "*maybe*" as a politeness disagreement hedge were also found in this context. 1 occurrence of "*maybe*" was found in a context of justifying personal preferences.

In conversation (7), subject 7 is justifying why she would like to get to know the North of Brazil:

(7)

<S1> … which area of the country would you like to get to know better?

<S7> <SE> pause</SE> Uh <SE> pause</SE> maybe the North.

<S1> Why?
<S7> Cause is so far and so different from here. It's hotter <SE> voice fading</SE> hottest, hotter, and I don't know is a lot of beach and the <SE> pause</SE> Amazon <SE> pause</SE> forest? **Maybe will be** nice.
(FCE P1- B1 BR-03)

Subject 7 initially makes a few assertions about the North of Brazil. She then produces another utterance initiated by "*maybe*" as an afterthought, summarising her attitude to the propositions previously expressed. "*Maybe*" functions as a mitigating device of the degree of assertiveness of her desire to visit the North of Brazil. It could be replaced by "*probably*" for the same purpose though it would imply a higher degree of assertiveness "*Probably it will be nice*". If replaced by the modal verb "*may*", it still retains a mitigating effect but seems to focus more on the degree of certainty of the future prediction "*It may be nice*".

(8)
<S8> Hum, **I'm not so sure**. Uh I think the music is the more attractive thing. This kind of offer, **I don't know**.
<S9> Yeah, but the uh, **I mean**, in the first uh the first time you, you go to a place, you go for music? **Maybe, maybe** you are but uh <SE> pause</SE>a free coffee is good too <SE> laughing</SE>!
(FCE P3- B1 BR-04)

Conversation (8) displays a number of hedging devices. Subjects are discussing the topic of a café. Subject 8 uses more explicit hedging forms to indicate his lack of certainty on the effectiveness of the suggestions proposed "*I'm not so sure*" and "*I don't know*". Subject 9 produces the backchannel "*yeah*", implying initial agreement[47], which is then immediately followed by the conjunction "*but*", indicating pseudo-agreement. "*Yeah, but…*" is then minimised by the implicit hedge "*I mean*". In her next utterance, Subject 9 produces 2 consecutive occurrences of "*maybe*" to soften disagreement with the idea that music is the most attractive thing.

[47] The function of "*yeah*" as a convergence response token will be addressed in chapter 6.1.1.

(9)

<S10> Uh **I think probably** depends of the place where this café will be. **Maybe the one free coffee** uh <SE> pause</SE> the cheaper, **I think it's not so**, it's, it's <SE> pause</SE> **will not be** <SE> pause</SE> a dif <SE> pause</SE> **will not be** a <SE> pause</SE>

<S11> Big difference.

<S10> Big difference uh to rich people.

<S11> Uhuh.

<S10> **Maybe the cheap** uh <SE> pause</SE> will be <SE> pause</SE> successful uh in uh poor parts of city, **maybe**. But if the place is good you, you can pay for that if your're <SE> pause</SE>, but not so expensive.

<S11> Yes, I agree.

(FCE P3- B1 BR-05)

Conversation (9) displays 3 occurrences of "*maybe*" and 1 occurrence of "*probably*". Subject 10 initially produces "*probably*" preceded by "*I think*" as a means to minimise the degree of certainty of his assertion. "*I think*" is clearly used as a fixed expression to introduce his opinion rather than as a hedging device. He then moves on to introduce a new suggestion: a free coffee. He marks his uncertainty about the effectiveness of this suggestion through his initial "*maybe*", which is strengthened by 3 consecutive negations. After 3 brief inserts, Subject 10 resumes his turn and produces 2 further occurrences of "*maybe*" as mitigating devices for expressing uncertainty, out of which 1 is in final position.

Let us now focus on the contexts of usage of both epistemic adverbs in medial position. Table 14 displays the distribution of "*probably*" and "*maybe*" among functions in the Brazilian corpus.

Table 14: Distribution of "*probably*" and "*maybe*" among functions in medial position in the Brazilian learner corpus

Function	Probably	Percentage	Maybe	Percentage
Mitigating the degree of certainty of propositions	9	75.00%	30	51.72%
Marking uncertainty when discussing pros and cons	3	25.00%	12	20.69%
Softening preferences	X	X	12	20.69%
Mitigating suggestions	X	X	2	3.45%
Planning ahead	X	X	2	3.45%
Total occurrences	**12**	**100.00%**	**58**	**100.00%**

Overall, the Brazilian learners of this study employed both "*probably*" and "*maybe*" in medial position mostly as mitigating devices to lower the degree of certainty of representative speech acts. The most frequent role of both forms was to lower the degree of certainty of their assertions. Most examples of this function were found in contexts in which individual learners were asked to compare situations shown in pictures, predicting what the people in the pictures were doing or thinking about and to draw some conclusions. Both forms were also employed as a means to express uncertainty vis-à-vis the effectiveness of the proposed suggestions during discussions. Nonetheless, the frequencies of both forms in each function varied.

As previously mentioned, "*maybe*" was much more frequent than "*probably*" in the Brazilian corpus. "*Maybe*" was also found to perform 3 further functions. It was employed in contexts in which learners were asked to justify their personal preferences and, thus, used as a means to soften the degree of assertiveness of the same. "*Maybe*" was also found within suggestions stems and as a hesitation device. Nonetheless, these two last functions were much less frequent if compared to the ones related to certainty and assertiveness. Thus, one can assume that both "*maybe*" and "*probably*" were employed in medial position mostly as more explicit pragmatic force modifiers of epistemic stance by the Brazilian learners of this study.

Let us now examine some of the above functions. In conversation (10), "*probably*" is produced towards the end of a long turn:

(10)

<S1> Thank you. Now … here are your photographs. They show people of different ages on educational visits <SE> pause</SE>. I'd like you to compare the photographs and say what you think the people will learn on their visits.

<S21> Uh.

<S1> All right?

<S21> Okay. In the, the first photograph I can see there are more old people, people with more age, they are, they are learning about history, they are visiting a museum and they are, you can see that, of course, there are people with more age, they are taking notes about the, about the photograph, about the, about the history, you can see they are inside a museum, and you can see they are very interested in the, in those things that they are looking for. And in the second picture, you can, you can see there are a , **maybe** a, a school , a school <SE> pause</SE>, a group of students or , **I don't know**, a group of friends, they are visiting a, visiting a park and they are looking for the, the nature, they are not, they are not looking for the history, they are looking more for their, they are learning about the nature, they are looking for the water, the plants, the trees, they are, **I think they probably will list**, will learn things more, more related to, to the nature . On the other hand, the first picture they are learning more, things more related to history.

(FCE P2- B1 BR-010)

The above conversation (10) shows a very confident answer to the proposed task. Subject 21 produces the above long turn within approximately 1 minute without much hesitation. There are a number of occurrences of "*I*/*you can see*" as a means to indicate sensory evidence. Subject 21 produces the hedges "*maybe*", "*I don't know*" and "*probably*" within his turn to lower the degree of certainty of his assertions. "*Probably*" is prefaced by "*I think they*" and followed by "*will*" implying a future prediction.

(11)

<S1> Uh do you plan your holidays carefully or do you prefer **just** to go?

<S10> Uh it depends the time I have to spend. If it uh is a big holidays uh <SE> pause</SE>so I can <SE> pause</SE> I prefer to plan it, **maybe** talking to my friends and invite and <SE> pause</SE> inviting them to <SE> pause</SE> to go to a, to the beach, for example. Uh but it's a holiday, a little holiday, I prefer **just** go out to visit my family.

<S1>Uh … tell us about a TV programme you've seen recently.
(FCE P1- B1 BR-05)

In the above conversation (11), subject 10 gives an extended answer to the proposed question, which is about personal preferences. One can notice two occurrences of the particle "*just*" paraphrasable as "*simply*" (functions to be detailed in the next sub-section). Subject 10 uses "*maybe*" as a means to make his preference less assertive or categorical.

(12)
<S20> Uh **if** it is open in the night<SE> pause </SE**>** **I don't know**, I, **I think it's not really** +
<S21> <SE> overlapping </SE> Yeah, **not sure if** it's a good idea also.
<S20>+ good because uh a coffee **maybe if**, if it sells more than coffee but just coffee at night, **I don't know if**+
<S21> <SE> overlapping </SE> Yeah +
<S20> +**if** it would be a good, good idea.
(FCE P3- B1 BR-010)

In conversation (12), subjects make extensive use of hedging devices to mark their uncertainty vis-à-vis the effectiveness of the suggestions proposed. There are a number of "if-clauses" signalling uncertainty and doubt. The hedges "*I don't know*", "*not sure*" and "*maybe*" are inserted within the "if-clauses" making the discourse even less assertive.

(13)
<S18> Uh <SE> pause</SE> if you want to, to, to attract much, much, much people <SE> laughing </SE>in your coffee **maybe you can** put the television with the uh football game.
<S19> <SE> laughing </SE> Coffee with a video game.
<S18> <SE> chuckles</SE>. But **probably** the, the, the café will uh will <SE> pause </SE> will lost some characters of a coffee.
<S19> Uhuh <SE> chuckles</SE>.
(FCE P3- B1 BR-010)

In the above conversation (13), subjects are discussing the topic of a café. Subject 18 makes a suggestion, which is prefaced by the adverb "*maybe*" (as a mitigating device) and encoded in the stem "*you can*". Subject 18 produces "*probably*" in his next turn as a means to indicate his uncertainty about the effectiveness of his own suggestion. This conversation is marked by laughter and involvement.

(14)

<S1> <SE> laughing</SE> Which area of the country would you like to get to know better?
<S14> Hum <SE> pause</SE> uh <SE> long pause</SE> **maybe** <SE> pause</SE> I don't know. Can you ask me again? Sorry, I kind of
<S1> <SE> overlapping</SE> Which area of the country would you like to get to know better?
<S14> Hum, Paraná.
(FCE P1- B1 BR-07)

Conversation (14) displays one occurrence of "*maybe*" as a planning device. In other words, subject 14 does not really understand the previous question and signals lack of understanding though filled and unfilled pauses. "*Maybe*" fills one of these pauses. Subject 14 then initiates a repair sequence, which is successfully accomplished.

Now that the functions and contexts of usage in medial position have been discussed, let us address "*maybe*" and "*probably*" in utterance final position. As detailed in table 13, the Brazilian corpus displays 4 occurrences of "*maybe*" and 1 occurrence of "*probably*" in utterance final position. All 5 occurrences show the epistemic adverbs being used as mitigating devices of the degree of certainty of concluding remarks as shown in the two examples below.

(15)

<S20> I agree with you.
<S21> **In my opinion** the two most important are the, of course, the music and the free coffee, **probably**.
(FCE P3- B1 BR-010)

(16)

S1> Young people usually go to different places to relax than older people. Why do you think this is?

<S16> Uh I think because the interesting, the activity is really different . The young people go to a, a place more , have more uh persons, it's more <SE> pause </SE> like crowded and the old people want to relax and stay alone **maybe** uh reading and stay calm, I think.

<S17> At home, **maybe**.

(FCE P4- B1 BR-08)

In conversation (15), "*probably*" is produced at the end of the collaborative task as a mitigating device of the conclusion reached. The concluding remark is prefaced by another hedge *"In my opinion"*. In conversation (16), there are two occurrences of "*maybe*" functioning as a mitigating device of the degree of certainty of the assertions. However, the second occurrence mitigates the concluding remark and also marks the closure of the topic.

Now that we have presented the functions of both epistemic adverbs in utterance initial, medial and final positions, let us summarise these data. Table 15 displays the distribution of "*probably*" and "*maybe*" among functions in all positions in the Brazilian corpus.

Table 15: Distribution of "*probably*" and "*maybe*" among functions in the Brazilian learner corpus

Function	Probably	Percentage	Maybe	Percentage
Mitigating the degree of certainty of propositions	10	77.00%	34	47.89%
Marking uncertainty when discussing pros and cons	3	23.00%	19	26.77%
Softening preferences	X	x	12	16.91%
Mitigating suggestions	X	x	2	2.81%
Planning ahead	X	x	2	2.81%
Politeness disagreement	X	x	2	2.81%
Total occurrences	**13**	**100.00%**	**71**	**100.00%**

Taking all functions into account, one can conclude that "*maybe*" was the most frequent hedge used as an explicit pragmatic force modifier of epistemic stance by the Brazilian learners of this study. Epistemic stance was found in contexts of mitigating the degree of certainty of assertions and conclusions, of expressing uncertainty vis-à-vis the effectiveness of suggestions and of softening the degree of assertiveness of one's preferences. "*Maybe*" was also found to perform the functions of mitigating suggestions, as a time planner and as a politeness disagreement token, though with a considerably lower frequency. "*Probably*" was much less frequently used by the learners of this study and exclusively used to mark degrees of certainty of assertions and when discussing the pros and cons of ideas, as displayed in Table 15.

To conclude our discussion, let us compare the frequency of both epistemic adverbs with other learner and native speaker data. Table 16 displays the raw and the relative frequencies (per 10,000 words) of "*maybe*" and "*probably*" in the native speaker benchmark corpora, namely the BNC Sampler spoken sub-corpus and the face-to-face conversations from the 1990s, in the Brazilian learner corpus and in the Spanish component of LINDSEI.

Table 16: "*Maybe*"and "*probably*" frequencies in 4 corpora

	BNC		**F-2-F**		**Brazilian corpus**		**Spanish corpus**	
	Raw frequencies	Relative frequency	Raw frequencies	Relative frequency	Raw frequencies	Relative frequency	Raw frequencies	Relative frequency
Maybe	309	**2.7103**	98	**3.6114**	71	**29.8281**	162	**14.3463**
Probably	524	**4.5962**	218	**8.0336**	13	**5.4614**	39	**3.4537**

The comparisons between native speaker data and the Brazilian findings were already addressed in the previous section. If one compares Spanish learner results with the Brazilian data, it is possible to conclude that both groups of students produced "*maybe*" more frequently than "*probably*". However, individual frequencies varied. While the subjects of this study produced "*maybe*" almost six times more often than "*probably*", the Spanish learners displayed a lower difference rate. In addition, the modal verbs "*may*" and "*might*" were present in the Spanish production, 11 occurrences each (0.9741 relative frequency per 10,000 words). Although the relative frequency of both modals was not high, all occurrences

displayed the epistemic use of the modals. Nonetheless, epistemic "*may*" and "*might*" were not observed in the Brazilian learner corpus. This might suggest that as learners acquire new forms for expressing similar meanings, the over-reliance on particular forms may weaken.

For illustrative purposes, the English Profile findings[48] (Hawkins & Filipovic, 2012, pp. 131-133) indicate that the meanings of both modal verbs are acquired at different levels of learner proficiency:

a) CEFR A2: May/Might Possibility (Epistemic)
b) CEFR B1: May Permission (Deontic)
c) CEFR C1: Might Permission (Deontic)

All in all, it is possible to say that the subjects of this study and the Spanish learners of LINDSEI employed "*maybe"* more frequently than "*probably*", differing from British English linguistic behaviour and aligned with American English patterns (Biber et al., 1999). While the Spanish learners were also able to produce the modal verbs "*may*" and "*migh*t" to convey degrees of certainty, this command was not yet observed in the Brazilian data, also differing from the English Profile results.

In this section we have investigated the functions of the epistemic adverbs "*probably*" and "*maybe*" and detailed their positions and uses in the Brazilian learner corpora. The subjects of this study produced both epistemic adverbs mostly in medial position. The second most frequent position was "*maybe*" utterance initial. Both adverbs were also found in utterance final position with lower frequencies. "*Maybe*" was the most frequent hedge used as an explicit pragmatic force modifier of epistemic stance. Epistemic stance was found in contexts of mitigating the degree of certainty of assertions and conclusions, of expressing uncertainty vis-à-vis the effectiveness of suggestions and of softening the degree of assertiveness of one's preferences. "*Maybe*" was also found mitigating suggestions, as a time planner and as a politeness disagreement token, though with a considerably lower frequency. "*Probably*" was much less frequently produced and exclusively used to mark degrees of certainty of assertions and when discussing pros and cons. The subjects of the study and the Spanish learners of LINDSEI employed "*maybe"* more frequently than "*probably*" for epistemic stance, differing from British English linguistic behaviour and aligned with American English patterns (Biber

[48] The English Profile learner findings were addressed in section 2.4.2.

et. al, 1999). While the Spanish learners were also able to produce the modal verbs "*may*" and "*might*" to convey degrees of certainty, this command was not observed in the Brazilian data.

5.1.2. "Just"

This section initially details Aijmer´s (2002) categorisation of the functions of "*just*" and her corpus-based findings in a British English native speaker spoken corpus (The London-Lund corpus). It then presents Brazilian learner corpora data (position, frequency and functions), qualitative analyses and comparisons with other corpora. It also displays four-word lexical bundles retrieved from the benchmark corpora, namely the BNC Sampler spoken sub-corpus and the face-to-face conversations from the 1990s, so as to illustrate less frequently observed pattterns or collocations in the Brazilian data.

Aijmer (2002) investigates the functions of "*just*" as an interpersonal particle in the 500,000 word London-Lund Corpus. Aijmer (2002, p. 153) indicates that "*just*" is strongly favoured in speech "where it is oriented to the involvement of the discourse partners in the speech event and functions as a subjective or interpersonal modal particle". "*Just*" is polysemous in present-day English and it works both as an adverb and a discourse particle. "*Just*" is used as a restrictive adverb paraphrasable as "*exactly*" or "*only*" with propositional meaning. In the conversation below (1), the function of "*just*" is to focus on the phrase "*beyond Swinton***"** and it suggests that you leave the motorway minimally or marginally beyond Swinton (Aijmer, 2002, p. 155):

(1)

A> the M ^! f\/our# ---

B> ^I'leave it :south of R\eading# ---

A> Well ^Jo 'leaves it: **just be'yond :Sw\indon**# -

c> no no

B> ^n\o#.

c> *you leave it ((till you're at the end))*

B> ^oh*'for.for ^coming'up :h\ere* you'mean

"*Just*" also has a temporal meaning as in "*just now*". This meaning is illustrated in conversation (2), where "*just*" has the reading "in the sense that the contrary situation almost applies" (Aijmer, 2002, p. 155):

(2)

B> ^/\ah' yes#-
what does that ^do in t\ea#
does ^that diss\/olve in 'tea#
A>I've ^**only just** discovered 'that#
a ^week a!g\o#
B> ((we)) used to ^have that in the !w\ar#
A>I ^had it in !c\offee#.

Conversely, the discourse particle differs from the restrictive adverb as it signals involvement in the discourse event as in the conversation below (Aijmer, 2002, p. 155):

(3)

b> you got a ^c/old#
A> - "^n\o#.
just a ^bit ! sni\iffy#
cos I'm – I " ^\am c/old#
and I'll ^be all right 'once I've warmed\up# -

In the above conversation (3), in addition to signalling involvement, "*just*" seems to have also been employed as a hedging device to miminise the seriousness of speaker A's health problem.

Aijmer (2002) indicates that "*just*" can be used as an emphasiser and for planning ahead (e.g. filling a pause before speakers move on with their turn). Aijmer (2002) clarifies that the emphatic "*just*" occurs in expressive contexts with intensifying meanings. In (4), the immediate context is the adjective "*frightful*" (Aijmer, 2002, p.156):

(4)

A> The"^gh\astly th/ing#.
^was [dhi:] - !senior common-room conver:s\ation# ./
that ^one had to cond\uct#.

which was ^**just !fr\ightful**# -
^and the --- a"!\absolutely {gro^t\esque#}#
^ritual#
of ^high table formal !d\inner#

Aijmer (2002) details the distribution of the meanings of "*just*" in the London-Lund Corpus. The emphatic function (including downtoning and paraphrasable as "*simply*", "*absolutely*", "*really*", "*truly*", "*only*" and "*merely*") represents the unmarked use (70.69%) in speech. The second most common function is the particulariser "*just*" as "*exactly*" (12.91%) followed by the temporal one (11.42%) as "*only just*" or "*just about*". The planning function is the least frequent (04.96%).

As for the subtleties of meaning, Aijmer (2002) clarifies that "*just*" is an intensifying or modal discourse particle modifying a proposition. It has an indexical relation to the speakers' attitudes or emotions towards a discourse event. Thus, it is never semantically neutral but has an evaluative overlay. In (5), "*just*" has a weakening or downtoning function (Aijmer, 2002, p. 159):

(5)
A> "^how 'long did you :g\o for#
B> ^=ooh#
^this was. **'just a w\eekend**#
A> ^y\/eah#.
^[\mhm]#
B> that we ^had._actually on'Ilkey :M\oor#

Aijmer (2002) explains that the downtoning function distinguishes "*just*" from "*only*" (and from "s*imply*"). In conversation (5), "*just a weekend*" is not only restrictive in the sense of restricting the application of the utterance predominately to the part focused (a weekend) but it implies "not this much" or "only to a small extent". Furthermore, "*just*" may have a reassuring overtone. According to Wierzbicka (1991, p. 351), "*just*" lends itself to mildly positive (reassuring, defensive, apologetic, even praising) interpretations because while "*small*" can easily be viewed as unimportant, it can also be viewed as desirable.

"*Just*" is also suited to politeness and persuasion. When "*just*" is used as a downtoner, it may be oriented to negative politeness. As previously mentioned in section 1.2.2, negative politeness strategies centre on the non-interference with hearers' wants. In the below conversation (6), the speaker is describing what answers she gave in an interview. The topic is a very old illuminated manuscript (Aijmer, 2002, pp. 168-169):

(6)

A> and ^I kept !st\ressing#
you ^kn\ow#
I'm ^not offering you :kn\owledge#
I'm ^just [tra] ^**I just think** this is an interesting thing to expl\ore#-

In the above conversation, "*just*" collocates with "*I think*" and modifies the assertion in a similar way to other downtoners like "*as well*", "*I mean*", the past progressive, the tag question and "*I suppose*". Aijmer (2002) points out that assertions, questions, suggestions, criticisms and requests are face-threatening acts whose effect may not be welcome to hearers. "*Just*" is used in negative politeness as a way of avoiding or softening FTAs, for instance: "*I was just thinking...*", "*Could you just...*" and "*I'm just wondering…*".

Conversely, in positive politeness, "*just*" is emphatic and intended to increase common ground. As previously mentioned in section 1.2.2, positive politeness strategies focus on the positive self-image that interactants claim for themselves. Strategies rely on the indication that, in some respects, speakers want some of hearers' wants. Aijmer (2002) indicates that "*just*" as a positive politeness strategy occurs with special scalar adjectives and collocations (e.g. exclamations with "*so*" and "*such*"). When it expresses positive politeness, it frequently occurs in clusters of strong and weak modifiers.

Aijmer (2002, pp. 164-165) adds that when "*just*" refers to an unusual situation, a strong emotion, or action, it can be paraphrased by maximisers such as "*absolutely*", "*perfectly*" and "*utterly*". These adverbials are not scalar but are sensitive to whether the following adjective expresses a high degree of a property, for instance: *"...and the decorations are* ***just appalling***". "*Just*" also occurs before extreme verbs suggesting that speakers are exaggerating: "*…and I thought* ***I would just die*** *in this set up*".

Now that the functions of "*just*" as an adverb and as a pragmatic marker have been discussed, let us focus on the occurrences of "*just*" in the Brazilian data. The Brazilian learner corpus displays 20 occurrences of "*just*". 3 occurrences have been disregarded as they were produced within the interlocutor's frame. Table 17 summarises the distribution of "*just*" in the Brazilian corpus vis-à-vis its functions and positions. The percentage figures refer to the distribution of "*just*" among functions.

Table 17: Distribution of "*just*" among functions in the Brazilian learner corpus

Just	Immediately after a preface	Medial position	Total	Percentage
Emphatic (including downtoning)	X	9	9	52.94%
Restrictive adverb	X	7	7	41.18%
Conditional - just if	1	X	1	5.88%
Total occurrences	**1**	**16**	**17**	**100%**

The above results indicate that the Brazilian learners of this study produced "*just*" mostly in medial position. Only one occurrence of "*just*" was found following a preface ("*yes*") in initial position. Overall, "*just*" was more frequently used in medial position to emphasise elements within utterances as a means to modify the pragmatic force of utterances (mitigating). The second most common function was the use of "*just*" as a restrictive adverb, also in medial position, taking scope over the truth condition of particular elements within the utterance. Therefore, the empathic function of "*just*" seems to be the unmarked use by the Brazilian learners of this study, similarly to Aijmer's findings (2002) in native speaker corpora. Nonetheless, the percentage difference between the emphatic and restrictive functions in the Brazilian corpus is much lower, possibly signalling a stronger tendency to use "*just*" as a restrictive adverb as well.

The absence of "*just*" as a temporal adverb may indicate subjects' unfamiliarity with this function or fewer opportunities for its realisation during the interviews. The absence of "*just*" as a planning device suggests that the Brazilian learners of this study tend to employ other

forms to perform this function, as demonstrated in the qualitative analyses of chapters 4 and 6 (to be detailed), and may not necessarily associate " *just*" with this function. In addition, the effect of "*just*" on the pragmatic force of utterances produced by the Brazilian learners was sometimes fuzzy, even within its contexts. This corroborates Nikula's assumption (1996) that implicit pragmatic force modifiers such as "*just*" lend themselves to different interpretations and may remain ambiguous and fuzzy even in the context.

Let us now discuss some contexts of usage of the emphatic and restrictive functions of "*just*" and attempt to specify the effects of the emphatic function. In conversation (7), subject 12 has been asked a question that requires an extended answer:

(7)

<S1> I'd like you to compare the photographs and say why you think the music is important to the different groups of people. All right?

<S12> Yes <SE> clearing his throat</SE>, <SE> pause</SE> so uh I think music is a very good uh way to express, to express our culture, like here in Brazil we have Samba, I can see in those pictures, I think here, the first one is about Africa, is in Africa, Africa song, I don't know. Uh <SE> pause</SE> and the second one is , I think is more classic, different way but I can see those, those faces, they are happy, I think music, music, dance, é uh <SE> pause</SE> express uh a happy moment and I think it's a very good uh way to, to make like a party, or<SE> pause</SE> I don't know. **I just think** it's a <SE> pause</SE> it's a good way to, to express our, our happiest moments.

<S1> Okay, Thank you.

(FCE P2- B1 BR-06)

In the above conversation, subject 12 produces a long turn in reply to the interviewer's question. He first compares why the music is important to each of the groups of people shown in the two pictures. He then concludes his ideas by saying why music is important in general. The occurrence of "*just*" is placed between the pronoun "*I*" and the attitudinal verb "*think*" and performs an emphatic function: it mitigates the degree of assertiveness of the statement, making his opinion lesson categorical.

(8)

S1> …, how much time do you spend at home?

<S5> Uh, not much, much time. I spend all morning and all afternoon in work and uh **just during the night**. And if I don't have any English class or other activities so uh a little time.
(FCE P1- B1 BR-02)

In conversation (8), there is one occurrence of "*just*" which could be interpreted in different ways. Initially, as a restrictive adverb indicating that the speaker only spends time at home during the night. Nonetheless, it may also be viewed as an emphasiser strengthening the assertion that this is not much time. A third interpretation could be that the speaker uses "*just*" as a face-saving strategy in the sense that spending too much time at home could be viewed as undesirable by the interviewer. Therefore, he further mitigates his assertion by producing "*just*".

(9)
<S3> I <SE> pause </SE> I'd like to, to add about the <SE> pause </SE> the advertising and the money that, that have on, on television, especially on the, on the matches and it involves a lot of, of people and I think it's a passion at the world, in the world, so a lot of people watch and give money to another person or a group.
<S2> Yes and uh the places that where there's uh people uh playing music people go **not just to eat or watch games** but to dance too and uh maybe the bar or the restaurant that have this, maybe can have uh be part of the nightlife of some people, it could change.
(FCE P3- B1 BR-01)

In conversation (9), subjects have been discussing different ideas on how to make a café more popular. Subject 2 produces "*just*" after a negation. This use is emphatic in the sense that it disputes a proposition: people go to places that offer live music not only to eat and watch games but also to dance. Subject 2 marks her stance by using "*just*" and then justifies her opinion. Nonetheless, this extract is particularly problematic as the ideas expressed by both speakers do not seem to make sense together in the order they are presented. In other words, this extract seems to lack coherence.

(10)
<S1> …Would you like to spend time in a café like this?
<S13> Sure.

<S12> Sure.
<S1> Why?
<S13>Uh first of all because of cappuccino, mocaccino +
<S12> <SE> overlapping</SE> Yes.
<S13> + and **things like** that cause I am addicted to <SE> pause</SE> coffee and coffee shops. Uh I get addicted to that when I went to New York and I<SE> pause</SE> and I went for the first time in a, in a Starbucks.
<S12> Starbucks, yeah.
<S13> **I just get addicted to that**. I get out from a Starbucks and walk a block and then entered another one, that's the way I go so <SE> pause</SE> I really love a café.
<S12> Me too, same, same opinion, **I think**.
(FCE P4- B1 BR-06)

In conversation (10), subject 13 is justifying why he would like to work in a café. He initially mentions his appreciation for types of coffee and uses the expression "*things like that*" as a means to indicate vagueness. He then relates his addiction to coffee and coffee shops like "*Starbucks*". Subject 12 produces an example of "*yeah*" as a response token[49] preceded by the lexical repetition of Starbucks. In the next turn, subject 13 uses "*just*" in the emphatic function so as to strengthen the degree of irresistibility of coffee. Subject 12 produces the fixed expression "*I think*" to conclude his opinion rather than as a hedge.

(11)
<S17> What do you think, need to have a lot of languages in the menu+
<S16> <SE> overlapping</SE> Yeah!
<S17> + or **just English and Portuguese**?
<S16> Yeah, I think English, Portuguese and another Asian food uh Asian uh language like Chinese, is really important because it's not every Chinese know English, yeah?
< S17> And the, the other thing…
(FCE P3- B1 BR-08)

In the above conversation (11), we can observe one example of "*just*" as a restrictive adverb, taking scope over one constituent of the utterance, i.e. focusing specifically on "*English and*

[49] "*Yeah*" as a response token will be addressed in section 6.1.1.

Portuguese". In other words, the speaker restricts the options for the menu to English and Portuguese only.

(12)
<S11> I like this place, not a coffee so I prefer the sofa, <SE> pause</SE> yes.
<S10> The first about the musicians, I think, not <SE> pause</SE> and about the television, totally not uh <SE> pause</SE>
<S11> Necessary.
<S10> Who gonna drink a coffee watching football?
<S11> Yes**, just if** he don't have TV at home.
<S10> Yeah <SE> laughing</SE>.
(FCE P3- B1 BR-05)

Conversation (12) displays the single occurrence of "*just*" in utterance initial position, preceded by the preface "*Yes*". "*Just*" is followed by a conditional clause and can be paraphrased as "*only if*". It may also be seen as a restrictive adverb in the sense that only people who do not have a TV at home will drink coffee while watching football.

(13)
<S1>Uh <SE> pause</SE> have you got any plans for this weekend?
<S12> Uh this, study. I have so much, so many tests next, next, in this month and the next week so I have to stay at home studying but my father and my mother now he's, they are with me and with my brother, they don't live here in Porto Alegre, they live in Espirito Santos and sometimes in Caxias do Sul, but in this, this weekend they stay here with us and I think **they are just going** to shopping uh have meal together but I, but I really need to study.
<S1> Okay… uh tell us about a TV programme you've seen recently.
(FCE P1- B1 BR-06)

An initial interpretation of "*just*" in the above conversation (13) would indicate its use as a restrictive adverb: it specifies the particular activities that the subject's parents are going to do together, i.e. they are exactly going shopping and they are having a meal together. Nonetheless, a closer look at the context of this long turn indicates that subject 12 is justifying

the reason why he will not be joining his parents as he needs to study for tests. Thus, "*just*" marks his stance and has a downtoning function, in the sense of less than ideal.

Similarly to Aijmer's (2002) findings, the analyses of the occurrences of "*just*" in the Brazilian data also attest that "*just*" is a particularly ambiguous pragmatic modifier device as it allows for different interpretations within the same context. It is also more oriented to the addressee of the message rather than to the message itself. Thus, it seems appropriate to place it towards the implicit end of the explicit-implicit continuum:

a) The Brazilian learners of this study were able to produce "*just*" in its emphatic function signalling involvement in the discourse;

b) Examples of the emphatic function included "*just*" as "*simply*", and both as mitigating and strengthening devices;

c) Some examples of "*just"* as a restrictive adverb seem to allow for further pragmatic interpretations, which highlights the high context- sensitivity of the pragmatic force modifier "*just*".

Now that we have illustrated the functions of "*just*" in the Brazilian data, let us compare frequencies with other learner and native speaker corpora. Table 18 displays the raw and the relative frequencies of "*just*" in utterance initial and in other positions (medial and final positions) in 4 corpora: the BNC Sampler spoken sub-corpus, the face-to-face conversations from the 1990s, the Brazilian learner corpus and the Spanish component of LINDSEI. All occurrences of "*just*" have been included irrespective of adverbial and pragmatic marking functions. Relative frequencies have been normalised to 10,000 words.

Table 18: Raw and relative frequencies of "*just*" in 4 corpora

	BNC		F-2-F		Brazilian learners		Spanish learners	
Just	Raw frequencies	Relative frequency	Raw frequencies	Relative frequency	Raw frequencies	Relative frequency	Raw frequencies	Relative frequency
Utterance Initial	316	2.7717	108	3.9799	1	0.4201	11	0.9741
Other position	3,568	31.2967	1,060	39.0625	16	6.7218	262	23.2020
Total	3,884	**34.0685**	1,168	**43.0424**	17	**7.1419**	273	**24.1761**

As both Brazilian and Spanish learner corpora include relatively small numbers of occurrences of "*just*", it was feasible to conduct manual investigations of their utterance positions. In the Spanish compontent of LINDSEI, only 6 occurrences were found in utterance final position, 4 of which were interrupted turns with overlapping talk. Therefore, one can infer that "*just*" medial was the most frequent position produced by both nationality types.

As for native speaker data, the CQP web server allows users to conduct case-sensitive searches. Thus, utterance initial figures refer to "*just*" in upper case while other position ones relate to its occurrences in lower case. Thinned searches of "*just*" lower case (100 occurrences) were then carried out in the BNC Sampler spoken sub-corpus. It was concluded that 99 occurrences displayed its use in medial position while only one single occurrence of "*just*" was found in utterance final. Thus, one can infer that "*just*" is much more frequently used in medial position by British native speakers, and it is 10 times more frequent than in utterance initial position.

By comparing all data from table 18, one can observe that both native speakers and learners produced "*just*" mostly in medial position. However, individual frequencies varied. The specialised face-to-face conversation displayed the highest frequency, strengthening the association of "*just*" with face-to-face conversations. The Spanish learners of LINDSEI were able to produce "*just*" in medial position with a much higher frequency than that of the subjects of this study. These results seem to indicate underuse of "*just*" in medial position by the Brazilian learners of this study. In addition, both Brazilian and Spanish learners displayed a low relative frequency of "*just*" in utterance initial position, which may suggest less familiarity with "*just*" as a politeness strategy to minimise the degree of imposition of face threatenings acts, especially when preceding imperative forms.

Let us now focus on the most common collocating items with "*just*" in medial position in native speaker production. Table 19 indicates the 10 most frequent four-word clusters in which the particle "*just*" constitutes the third element. The percentage refers to the distribution of the specific collocation in comparison to all observed collocations.

Table 19: Most frequent four-word clusters with "*just*" as a third element in the benchmark corpora

BNC			F-2F		
Collocation	Raw frequencies	Percentage	Collocation	Raw frequencies	Percentage
it's just a	33	0.92%	it's just a	14	1.32%
it's just that	26	0.73%	It's just the	9	0.85%
it's just the	21	0.59%	It's just that	7	0.66%
it's just,	12	0.34%	I'm just going	6	3.70%
I'd just like	12	0.34%	it's just like	6	3.70%
it's just like	10	0.28%	I was just going	5	3.70%
Can I just ask	9	0.25%	I'm just looking	4	2.78%
's not just a	8	0.22%	I was just saying	4	2.78%
I'm just going	8	0.22%	I was just thinking	4	2.78%
I've just got	8	0.22%	I'm just trying	4	1.85%

In both native speaker corpora, "*just*" is inserted in contexts of "*it is*" + "*just*" + a function word and "*I*" + auxiliary verb + "*just*". "*It is just…*" is a typical preface to justifications. The face-to-face conversation sub-corpus displays a far greater frequency of combinations of "*I*" + "auxiliary verb" + "*just*" in the progressive aspect, possibly signalling a tendency to mitigation, indirectness and negative politeness. None of the above clusters were present in the Brazilian production. The Brazilian learner corpus displayed only 3 occurrences of "*just*" as a third item within a four-word cluster: "*they are **just** going*", "*are not **just** reading*" and "*there is **just** one*". There were 3 occurrences of "*just*" as a second element preceding verbs "*I **just** think*", "*I **just** get addicted"* and "*we **just** met*". The remaining occurrences displayed *"just"* in medial position modifying specific utterance constituents. One occurrence followed a preface in utterance initial position.

Concluding, in this section we have investigated the position and the functions of "*just*" in the Brazilian learner corpus and compared results with native speaker and Spanish learner data (LINDSEI). "*Just*" works both as an adverb and a discourse particle (Aijmer, 2002). It is used

as a restrictive adverb paraphrasable as "*exactly*" or "*only*" with propositional meaning while as a discourse particle, it signals involvement in the discourse event. The Brazilian learners of this study produced "*just*" mostly in medial position to emphasise elements within utterances, as a means to modify the pragmatic force of utterances (mitigating). The second most common function was the use of "*just*" as a restrictive adverb, also in medial position, taking scope over the truth condition of particular elements within the utterance. The analyses of the occurrences of "*just*" in the Brazilian data also attested that "*just*" is a particularly ambiguous pragmatic modifier and may be placed towards the implicit end of the explicit-implicit continuum. The comparison between benchmark and learner corpora indicated that both native speakers and learners produced "*just*" mostly in medial position. However, individual frequencies varied. Results indicated underuse of "*just*" in medial position by the Brazilian learners of this study. In addition, both Brazilian and Spanish learners displayed a low relative frequency of "*just*" in utterance initial position, which may suggest less familiarity with "*just*" as a politeness strategy to minimise the degree of imposition of face threatenings acts, especially when preceding imperative forms.

5.2 DISCUSSION AND RECOMMENDATIONS

Up to now, we have presented corpus-based findings vis-à-vis the pragmatic functions and relative frequencies of the adverbial hedging forms "*maybe*", "*probably*" and "*just*", based on British English native speaker spoken data and Brazilian and Spanish learner data. Let us now discuss these findings in the light of the specific research sub-question addressed in this chapter.

RSQ2: *What are the most common explicit and implicit adverbial hedges used to mitigate representative speech acts by Brazilian learners at CEFR B1? How do their frequency of use and pragmatic functions compare to those of native speakers?*

As shown in table 11 (section 5.1), the Brazilian learner corpus displayed 18 lexical items (or lemmas) which seem to fit the form characterisation of hedges as detailed by O'Keeffe, McCarthy and Carter (2007, p.175) and/or are traditionally associated with mitigation. Items with fewer than 4 occurrences were disregarded. The adverbs "*well*", "*really*" and "*actually*" were not considered since they had been analysed as discourse marking adverbs used to

mediate segments of discourse in chapter 4. The relative frequencies of items were as follows (per 10,000 words) : "*think*" (102.9282), "*can*" (56.2954), "*things/thing*" (34.4494), "*maybe*" (29.8281), "*about*" (26.4672), "*would*" (22.2661), "*will*" (19.3252), "*kind*" (17.6448), "*just*" (7.1419), "*seems/seem*" (7.1419), "*probably*" (5.4614), "*feel*" (4.2011), "*could*" (3.7810), "*(in) my opinion*" (3.7810), "*guess*" (3.3609), "*bit*" (2.5206), "*should*" (2.1005) and "*types/type*" (1.6804). It was concluded that the Brazilian learners of this study produced a range of word forms traditionally associated with mitigation: attitudinal verbs, modal verbs, nouns and epistemic and restrictive adverbs. Nevertheless, not all forms were necessarily employed for mitigating purposes.

The adverbial hedging forms "*maybe*", "*probably*" and "*just*" were found mostly within contexts of representative speech acts such as asserting, making deductions and concluding. They were chosen for the empirical investigations. Qualitative analyses corroborated the following assumptions:

a) As part of the class of adverbs, "*maybe*", "*probably*" and "*just*" can be placed in different utterance positions, displaying more versatility than attitudinal verbs and modal verbs, which are necessarily attached to preceding subjects. Thus, more varied uses were observed;

b) The selected adverbial hedges were employed for mitigating the degree of assertiveness and/or to lower the degree of certainty in contexts of assertions, deductions and conclusions;

c) "*Maybe*" and "*probably*" are more explicit forms of hedges while the pragmatic marker "*just*" is more ambiguous, lending itself to different interpretations;

d) The Brazilian corpus did not display a single occurrence of the modal verbs "*may*" or "*might*", which suggests a tendency to use adverbial equivalents to convey this degree of certainty instead of modal verb forms.

Let us now examine the distribution of the forms under investigation per subject. Table 20 displays the raw frequencies of specific adverbial hedges as produced by individual learners.

Table 20: Distribution of adverbial hedges per subject

Brazilian learners	Maybe	Probably	Just
Subject 2	8		2
Subject 3	3		
Subject 4			
Subject 5	3		2
Subject 6	5	1	1
Subject 7	9		
Subject 8			
Subject 9	7		
Subject 10	10	3	
Subject 11		3	1
Subject 12			6
Subject 13	3		2
Subject 14	1		
Subject 15			1
Subject 16	4		
Subject 17	2		1
Subject 18	6	3	
Subject 19	1		
Subject 20	6	1	1
Subject 21	3	2	
Total	**71**	**13**	**17**

The data above allow us to conclude that 80% of the Brazilian learners of this study employed at least one explicit adverbial hedge to mark epistemic stance. While 75% of the subjects produced "*maybe*", "*probably*" occurred in 30% of individual learners' production. 40% of the subjects clearly overused "*maybe*" (4 or more occurrences). In addition, 25% of the Brazilian learners employed two forms of explicit adverbial hedges to convey epistemic stance. As for the distribution of "*just*", 45% of the subjects of this study produced it either as an implicit adverbial hedge or as a restrictive adverb. Overall, only 25% of the subjects employed one form of explicit adverbial hedges and "*just*" (irrespective of functions). 10% of the Brazilian learners were able to produce the three adverbial hedging forms under investigation. One can conclude that most subjects of this study were able to employ explicit adverbial forms to hedge representative speech acts while "*just*" as an implicit modifier or as a restrictive adverb posed more problems to learners.

Let us now summarise frequency results related to specific forms. The Brazilian learner corpus displayed 71 occurrences of "*maybe*" and 13 occurrences of "*probably*". The comparison of relative frequencies between the native speaker benchmark corpora and the Brazilian learner corpus indicated that "*probably*" occurred more frequently than "*maybe*" in both British English native speaker corpora. The Brazilian learners of this study used "*maybe*" more frequently than "*probably*", differing from British native speaker linguistic behaviour and aligned with American English patterns (Biber et al., 1999). Relative frequencies (per 10,000 words) were as follows:

a) BNC Sampler spoken sub-corpus: "*probably*" (4.5962) and "*maybe*" (2.7103)

b) Face-to-face conversations from the 1990s: "*probably*" (8.0336) and "*maybe*" (3.6114)

c) Brazilian learner corpus: "*probably*" (5.4614) and "*maybe*" (29.8281)

Different reasons may account for the favouritism towards the epistemic adverb "*maybe*" by the Brazilian learners of this study. Brazilian learners in general have much more exposure to the American variety due to various reasons: Brazil is geographically closer to the USA, which may facilitate more frequent visits to the USA; most Brazilian schools (both regular schools and language institutions) adopt American English textbooks; a regular and more frequent access to American movies and sitcoms. Furthermore, the Brazilian learners of this study did not produce any occurrences of "*may*" and "*might*" for expressing degrees of certainty. This indicates that the subjects of this study resorted to the epistemic adverb "*maybe*" (and possibly *"probably*" with a much lower frequency) to perform this function. Lastly, the high frequency of "*maybe*" may be also due to transfer as in Brazilian Portuguese, epistemic stance is often conveyed through the adverb "*talvez*", the direct translation of "*maybe*".

The qualitative analyses of this chapter indicated that the subjects of this study employed "*probably*" and "*maybe*" to perform the following functions, as highlighted in table 15 (section 5.1.1):

1. "*Probably*"

a) Mitigating the degree of certainty of propositions – 77.00 %

b) Marking uncertainty when discussing pros and cons – 23.00%

2. "*Maybe*"

a) Mitigating the degree of certainty of propositions – 47.89%

b) Marking uncertainty when discussing pros and cons – 26.77%

c) Softening preferences – 16.91%

d) Mitigating suggestions- 2.81%

e) Planning ahead – 2.81%

f) Politeness disagreement – 2.81%

Considering the above data, it is possible to conclude that both epistemic adverbs "*probably*" and "*maybe*" were employed as explicit hedges to mitigate representative speech acts. All occurrences of "*probably*" displayed its use as an explicit adverbial hedge used to mitigate assertions, either mitigating the degree of certainty of propositions (when speculating about visual prompts) or marking uncertainty about the effectiveness of proposed ideas (reaching an agreement) . 94.38 % of the occurrences of "*maybe*" also showed its use as an explicit adverbial hedge employed to mitigate assertions, mitigating the degree of certainty of propositions, marking uncertainty when discussing pros and cons of suggested ideas, softening individual preferences and as a politeness hedge when expressing disagreement with proposed ideas. 2.81% of the occurrences of "*maybe*" demonstrated its use mitigating another speech act category, namely directives (suggestions).

As for the third item under investigation, the Brazilian learners of this study produced 17 occurrences of "*just*". Comparisons with native speaker data attested an extremely high frequency of "*just*" in both benchmark corpora and indicated low figures of "*just*" in the learner corpus, which signals underuse. It was also concluded that both native speakers and learners produced "*just*" mostly in medial position. Relative frequencies (per 10,000 words) were as follows:

a) BNC Sampler spoken sub-corpus: 34.0685

b) Face-to-face conversations from the 1990s: 43.0424

c) Brazilian learner corpus: 7.1419

As shown in table 17, the functions of the occurrences of "*just*" in the Brazilian learner corpus can be summarised as follows:

a) Emphatic (including downtoning) – 52. 94%

b) Restrictive adverb- 41.18%
c) Conditional – just if – 5.88%

"*Just*" was more frequently used to emphasise elements within utterances as a means to modify the pragmatic force of utterances (mitigating). Out of the 9 occurrences, 8 diplayed the use of "*just*" as an implicit hedge mitigating assertions. There was only 1 occurrence of "*just*" strengthening assertions. The effect of "*just*" on the pragmatic force of utterances produced by the Brazilian learners was sometimes fuzzy, even within its contexts. The second most common function was the use of "*just*" as a restrictive adverb, taking scope over the truth condition of particular elements within the utterance. Therefore, the empathic function of "*just*" seems to be the unmarked use by the Brazilian learners of this study, similarly to Aijmer's findings (2002) in native speaker corpora. Nonetheless, the percentage difference between the emphatic and restrictive functions in the Brazilian corpus was much lower, possibly signalling a stronger tendency to use "*just*" as a restrictive adverb as well.

Considering all learner data presented in this section, it is possible to conclude that the subjects of this study would benefit from pedagogical assistance for acquiring other forms to convey epistemic stance, based on the apparent overuse of "*maybe*" and the absence of "*may*" and "*might*" in the Brazilian data. Learners would also benefit from more exposure to the pragmatic functions of "*just*", due to relatively low figures observed in the Brazilian data. Within EFL contexts, the modal verbs "*must*", "*may*", "*might*", "*could*" and "*can't*" are usually grouped as "modal verbs for deductions" and introduced to learners in general as from CEFR A2. The pragmatic functions of "*just*", on the other hand, are usually presented to students at more advanced levels, usually as from CEFR B2 or C1.

Assuming that the subjects of this study will have already been exposed to the epistemic use of modal verbs in general, a reversed teaching cycle such as "practice, exposure and instruction" (Thornbury and Slade, 2006) might be a suitable option, as suggested for discourse markers in section 4.2. Resembling the FCE speaking task part 2, learners can be given some pictures that illustrate a specific topic such as friendship, work, leisure, etc. Learners then work in pairs and discuss the similiarities and differences among the pictures and how the people shown in the pictures may be feeling. Prior to this, teachers can record themselves performing the same task with another peer and ensuring they both employ

different forms to mark epistemic stance. After learners have carried out the activity, they can listen to its enhanced version, receive transcripts and underline forms used to express degrees of certainty. After discussing specific uses, learners can then repeat the same task (with different pictures illustrating another topic), consciously making an effort to employ the target forms.

Other practice activities that offer the opportunity to use different epistemic markers include exploiting unsual images or sounds for learners to draw conclusions as to what is going on, who the people are and what the people are doing there. Teachers can also write isolated pieces of information retrieved from a real story on the board. Learners then try to predict what the story is about by guessing the connection among the items. Alternatively, learners can also use their own images saved on their mobiles or tablets for other learners to make guesses about the situational contexts: where the people were, what they were doing there, what they did afterwards. This activity in particular would require the use of modal verbs with perfect infinitives, which is likely to be more challenging for learners.

In relation to "*just*", a traditional "exposure, instruction and practice cycle" (Thornbury and Slade, 2006) may be more effective since learners at CEFR B1 are likely to have been introduced to the uses of "*just*" only as a restrictive adverb, incluing the temporal meaning. As a corpus-based teaching suggestion, teachers can access a free online corpus such as COCA and the BNC (http://corpus.byu.edu/coca/) and carry out searches of "*just*". Once invididual meanings and functions are spotted, concordance lines can be downloaded or copied and pasted onto a word file. Teachers may also list the meanings or functions in another word file. In class, learners initially receive a copy of the concordance lines and predict the meanings of "*just*". Then, they receive the second file and match examples to functions. As a practice activity for the pragmatic uses of "*just*" , teachers can select a real life situation in which speakers are likely to use "*just*" in its emphatic use (mitigating and strengthening) and turn it into a roleplay like "teenagers asking permission to parents for something the latter are likely to say *no* to".

Now that we have proposed some pedagogical recommendations based on the findings of this chapter, let us comment on one result in particular. The apparent overuse of "*maybe*" in the Brazilian data offers a rich opportunity for further investigations within Interlanguage

Pragmatics. Is the observed tendency to use explicit adverbial hedges to convey epistemic stance instead of modal verb forms due to the fact that learners have not yet acquired the epistemic use of "*may*" and "*might*"? Could this tendency be also interpreted as negative transfer? Longitudinal learner investigations combined with cross-cultural corpus-based studies may provide some insight into this apparent overuse of "*maybe*" in the Brazilian learner corpus.

In this section we have summarised corpus-based findings vis-à-vis the relative frequencies and the pragmatic functions of the adverbial hedges "*probably*", "*maybe*" and "*just"* in the Brazilian learner corpus and in native speaker corpora. We have discussed these findings in the light of the specific research sub-question addressed in this chapter. We have also proposed some pedagogical recommendations based on observed tendencies.

SUMMARY

This chapter has investigated the most common explicit and implicit adverbial hedges used to mitigate representative speech acts by Brazilian learners at CEFR B1. Explicit modifiers include attitudinal adverbs, parenthetical constructions or adverbs of degree while implicit modifiers have a tendency to remain ambiguous or fuzzy even in the context, often making various interpretations possible (Nikula, 1996). Native speaker corpus evidence reveals different forms used for mitigating purposes (O'Keeffe, McCarthy and Carter (2007: 175): modal verbs and verbs with modal meaning, nouns, degree and stance adverbs, syntactic choices, double negative, evaluation relative clause insertion and features of "onlineness" adjustments. The 10 most frequent hedging items in the Limerick Corpus of Irish English were also reported: "*just*", "*really*", "*actually*", "*probably*", "*I think*", "*a bit*", "*kind of*", "*sort of*", "*you know*" and "*I suppose*".

AntConc's word list tool was used to view the rank positions and the raw frequencies of all the words in the Brazilian learner corpus. A manual list was made of all the lexical items which fit into the form categorisation of hedges and/or are traditionally associated with mitigation. Overall, the subjects of this study produced attitudinal verbs, modal verbs, nouns and epistemic and restrictive adverbs. However, not all selected forms were necessarily used for mitigating purposes. Analyses of concordance lines of individual items were carried out to

investigate common uses. The attitudinal verb "*think*" was the most frequently form associated with mitigation produced by the subjects of this study. However, this formula is widely used by Brazilian learners in general as a means to introduce their opinion rather than as a hedging device. The second most frequent form, the modal verb "*can*", was found mostly in contexts of suggestions, expressing possibility and ability. The third most frequent form was the lemma "*things/ thing*", used for expressing vagueness within categories. The adverbial forms "*maybe*", "*probably*" and "*just*" were found mostly within contexts of representative speech acts and were chosen for the empirical investigations of chapter 5.

Native speaker investigations attested an extremely high frequency of "*just*" in both benchmark corpora. As for the epistemic adverbs, "*probably*" occured twice as often as *maybe*" in the specialised spoken sub-corpus. It was also more frequent than "*maybe*" in the BNC Sampler spoken sub-corpus. By comparing native speaker results with the Brazilian data, it was suggested that "*maybe*" was possibly overused by the subjects of this study while the occurrences of "*just*" in the Brazilian corpus may indicate some possible underuse. Another salient feature was that the Brazilian learners of this study used "*maybe*" more frequently than "*probably*", differing from British native speaker linguistic behaviour and aligned with American English patterns (Biber et al., 1999).

Chapter sub-sections investigated the pragmatic functions and frequencies of the items under investigation in native speaker and learner corpora.The Brazilian data revealed that the subjects of this study produced both "*probably*" and "*maybe*" mostly in medial position. The second most frequent position was "*maybe*" utterance initial. Both adverbs were also found in utterance final position with lower frequencies.Comparisons with Spanish learner data (LINDSEI) indicated that both nationalities produced "*maybe*" more frequently than "*probably*". The subjects of this study employed "*maybe*" almost six times more often than "*probably*" whereas the Spanish learners displayed a lower difference rate. In addition, while the modal verbs "*may*" and "*might*" were present in the Spanish production, they were not observed in the Brazilian learner corpus. This might suggest that as learners acquire new forms for expressing similar meanings, the over-reliance on particular items may weaken.

The Brazilian learners of this study produced "*just*" largely in medial position and more frequently to emphasise elements within utterances as a means to modify the pragmatic force

of utterances (mitigating). The second most common function was its use as a restrictive adverb, also in medial position. It was concluded that the empathic function of "*just*" seems to be the unmarked use by the subjects of this study, similarly to Aijmer's (2002) native speaker findings. Comparisons with other learner corpora indicated the Spanish learners of LINDSEI were able to produce "*just*" in medial position with a much higher frequency than that observed in the Brazilian corpus. These results seem to indicate underuse of "*just*" in medial position by the subjects of this study. In addition, both Brazilian and Spanish learners displayed a low relative frequency of "*just*" in utterance initial position, which may suggest less familiarity with "*just*" as a politeness strategy to minimise the degree of imposition of face threatenings acts, especially when preceding imperative forms.

The final part of this chapter discussed "*probably*", "*maybe*" and "*just*" findings in the light of its research sub-question. It was concluded that both epistemic adverbs "*probably*" and "*maybe*" were employed as explicit hedges to mitigate representative speech acts. All occurrences of "*probably*" displayed its use as an explicit adverbial hedge used to mitigate assertions. 94.44% of the occurrences of "*maybe*" also showed its use as an explicit adverbial hedge used to mitigate assertions and as a politeness hedge when expressing disagreement with proposed ideas. It was concluded that "*just*" was more frequently used to emphasise elements within utterances as a means to mitigate the pragmatic force of utterances. The effect of "*just*" on the pragmatic force of utterances produced by the Brazilian learners was sometimes fuzzy, even within its contexts, corroborating its nature as an implicit modifier.

As for pedagogical recommendations, it was argued that the subjects of this study would benefit from pedagogical assistance for acquiring other forms to convey epistemic stance, based on the overuse of "*maybe*" and on the absence of "*may*" and "*might*" in the Brazilian data. Learners would also benefit from more exposure to the pragmatic functions of "*just*", due to the relatively low figures observed in the Brazilian data. Teaching suggestions were then detailed based on the three-part cycle "exposure, instruction and practice" (Thornbury and Slade, 2006), not necessarily in this order.

Let us now focus on the third and final facet of conversational competence, namely the deployment of conversational practices, by investigating minimal response tokens employed to express good listenership.

6. EXPRESSING GOOD LISTENERSHIP: MINIMAL RESPONSE TOKENS

The term "good listenership" was coined by McCarthy (2002) when addressing a set of words that routinely perform interactional/relational responsive functions. McCarthy states that Sinclair and Coulthard's (1975) tripartite classification of moves (addressed in section 1.1.1) is extremely useful to the study of spoken corpora as "the sequential positioning of words within the initiation -> response -> feedback model says a great deal about their typical environments of occurrence and their associated conversational functions" (2002, p. 51). As a result, the interpretation of a word may be affected not only by its syntactic and lexical properties but also by the contexts in which it typically occurs (e.g. as a one- word insert sequence). Good listenership requires more than just acknowledgement and transactional efficiency since listeners orient themselves towards the creation and maintenance of sociability and affective well-being in their responses (McCarthy, 2002).

Backchannels[50] are forms employed to express good listenership. They are "verbal and non-verbal devices used to provide feedback and other supportive responses, normally as a way to encourage the speaker to continue" (Carter & McCarthy, 2006, p. 892). Typical back-channelling devices include vocalisations, words and phrases such as "*mm*", "*uhum*", "*yeah*" and "*right*". Adolphs (2008, p. 121) explains that, in conversations, listeners produce verbal and non-verbal backchannels indicating that they are following the speaker's turns. These devices are usually produced when a speaker is talking at length. According to Adolphs (2008), there is evidence that the speaker relies upon backchannels as a means to interpret how the message is being received and that listeners respond with precision at appropriate moments and in appropriate ways to incoming talk. Adolphs (2008) clarifies that backchannels are not aimed at taking control of the floor but rather at offering relevant feedback to the speaker as they invoke different communicative choices.

Within their basic functions, Adolphs (2008, p. 121) states that backchannels can signal acknowledgement and agreement as the talk unfolds, mark newsworthy items, declare appreciation of a speaker's talk and clarify the sense and meaning of an utterance. They may also include laughter, sighs and other sympathetic messages that indicate hearers' active listenership. Head-nods, for instance, adopt the same conventionalised form on a basic level,

[50] The notion of "backchannels" (Yngve, 1970) was addressed in section 1.3.1 within CA domains.

namely the up and down motion of the head. However, similarly to verbal backchannels, they express more than one function. The meaning of head-nods is situation-dependent, likewise the meaning of vocalisations and other bodily gestures. I have narrowed the investigations of this chapter to verbal backchannels as the transcription of the Brazilian corpus does not indicate bodily gestures. In addition, I have chosen to analyse minimal forms (as opposed to non-minimal ones) so as to avoid overlap with the discourse marking adverbs and the adverbial hedges presented in the previous chapters.

This chapter addresses the following research question, focusing specifically on its sub-question:

RQ3: *How do Brazilian learners at CEFR B1 express good listenership?*
RSQ3: *What are the most common minimal response tokens used to express good listenership by Brazilian learners at CEFR B1? How do their frequency of use and pragmatic functions compare to those of native speakers?*

In the first section of this chapter, I introduce the distinction between minimal and non-minimal response tokens (O'Keeffe & Adolphs, 2008). I then present corpus-based findings vis-à-vis the distribution of forms and functions of response tokens in different varieties of English (O'Keeffe & Adolphs, 2008, Carter &McCarthy, 2006, McCarthy, 2002). Next I introduce the possible candidates for minimal response tokens in the Brazilian learner corpus and justify choices of items under investigation, namely the minimal forms "*yeah*" and "*uhuh*".

In the sub-sections of this chapter, I detail the contexts of usage of the selected minimal forms. I specify their distribution among functions as response tokens and illustrate their uses through qualitative analyses. I also compare relative frequencies with benchmark native speaker and Spanish learner data (LINDSEI).

In the final part of this chapter, I summarise the main points from the previous parts, discuss findings in the light of the proposed research sub-question and make some pedagogical recommendations.

6.1. FORMS AND FUNCTIONS OF RESPONSE TOKENS

In this section, I introduce the distinction between minimal and non-minimal response tokens and summarise corpus-based evidence on the forms and functions of response tokens (O'Keeffe & Adolphs, 2008, Carter &McCarthy, 2006, McCarthy, 2002). I then present candidates for minimal forms in the Brazilian learner corpus and justify choices for empirical investigations.

O'Keeffe and Adolphs (2008) investigate the distribution of forms and functions of response tokens in two varieties of English, British English and Irish English, using two native speaker corpora: The Cambridge and Nottingham Corpus of Discourse in English (CANCODE) and The Limerick Corpus of Irish English (LCIE). According to O'Keeffe and Adolphs (2008), the existing research categorises response tokens as "minimal" and "non-minimal" forms (e.g. Schegloff, 1982, Tottie 1991, McCarthy 2002). "Minimal response tokens" include short utterances like "*yeah*" and non-word vocalisations such as "*mm"* and "*umhum*" (O'Keeffe & Adolphs, 2008, p. 74). In addition, they are minimal forms as they have no homonyms in other word classes. "Non-minimal response tokens", on the other hand, comprise adverbs or adjectives functioning as pragmatic markers such as "*good"* and " *really"* and short phrases or minimal clauses such as *"you're not serious", "by all means"* and *" fair enough"*(O'Keeffe & Adolphs, 2008, p. 74). Nonetheless, the authors clarify that this division is not necessarily clear cut, especially when using a corpus of transcribed audio-recordings as these do not usually include head-nods and shoulder shrugs.

The below examples (1) and (2) illustrate the distinction between minimal and non-minimal forms, respectively (O'Keeffe & Adolphs, 2008, p. 75):

(1)

A: Tis is a lovely day but tis cold isn't it?

B: Ah the days are grand shure well yesterday was a bad bad evening.

A: **Mm**.

B: It turned black.

(LCIE)

(2)

A: …isn't that nice now. Blue sky.

B: **Lovely**.

A: A little breeze.

(LCIE)

As for co-occurring items, McCarthy (2002, p. 65) indicates that non-minimal response tokens may be pre-modified by intensifying adverbs, which add further emphasis such as "*most*":

(3)

[Discussing tenancy problems in rented accommodation]

A: Isn't there something in your tenancy agreement about it? You have a written agreement, don't you?

B: **Most definitely**.

(CANCODE)

McCarthy (2002) also states that both minimal and non-minimal response tokens can occur in pairs or clusters as in the example below (O'Keeffe & Adolphs, 2008, pp. 75-76):

(4)

A: …you know it reminds me of am the play and ah.

B: Mm.

A: And the character in the play is not +

B: I don't know.

A: +someone I'd kind of identify with+

B: Yeah that's true that's true but I wonder if that's a cultural sort of +

A: **Yeah mm.**

B: + I don't know I had the same question for Rosemary.

(LCIE)

Carter and McCarthy (2006) point out that response token pairings are particularly evident when a topic is being closed down or at a boundary in the talk when another topic is being introduced as in the examples below (O'Keeffe & Adolphs, 2008, p. 76):

(5)

[Couple asking permission to look at a disused railway line]

A: It went through, it goes through. Straight, straight on.

B: **Right. Wonderful. Great**. Can we look round then?

A: Yes, certainly.

B: Thank you.

(CANCODE)

With reference to the functions of response tokens, O'Keeffe and Adolphs (2008, p. 84) adopt the following framework for their classification in casual conversation:

a) **Continuer tokens:** they maintain the flow of the discourse. Typical examples include the minimal forms "*yeah*" and "*mm*";
b) **Convergence tokens:** they mark agreement and convergence. They are linked to points in the discourse where there is a topic boundary or where there is the need to converge on an understanding of what is common ground or shared knowledge between participants. Examples include single word items such as "*yeah*", follow-up questions such as "*Did You?"* and short statements, e.g. agreeing statements like "*Yeah, it's pretty sad*";
c) **Engagement tokens:** they mark high engagement where addressees respond on an affective level to the content of the message. They express genuine emotional responses such as surprise, shock, horror, sympathy, empathy and so forth. Examples include single-word forms such as "*excellent*", "*absolutely*", short statements and repetitions such as "*that's nice"*, "*oh wow"*, "*oh really"* and follow-up questions: "*did you*?";
d) **Information receipt tokens:** they mark points in the discourse where adequate information has been received. They can impose a boundary in the discourse and signal a point of topic transition or closure. They can even be indicative of asymmetrical discourse.

In order to investigate the distribution of tokens among functions, O'Keeffe and Adolphs (2008) compiled 2 small and comparative sub-corpora of 20,000 words each from LCIE and CANCODE. Each sub-corpus displays conversations of young women around the age of 20.

Qualitative analyses of response tokens indicated that British young women employed response tokens in the following order of frequency: convergence, engagement, continuer and information receipt. Convergence tokens occurred twice as frequently as engagement tokens, ten times more frequently than continuer tokens and 11 times more frequently than information receipt tokens (O'Keeffe and Adolphs, 2008, p.90).

Now let us address each of the above functions in more detail. Firstly, continuer response tokens are facilitative as they maintain the flow of talk by encouraging the speaker to continue. "Speakers perceive continuer response tokens as floor-yielding signals that mark the addressee's desire for the talk to continue" (O'Keeffe & Adolphs, 2008, p. 84). The below conversation (6) displays the minimal response form "*yeah*" surrounded by ongoing utterances. The use of "*yeah*" by the listener encourages the speaker to continue the story (O'Keeffe & Adolphs, 2008, p. 85):

(6)

A: And they sent one back saying "ah come on now Sinead are you messing or are you serious like?"

B: **Yeah**.

A: And ah he sent one saying "no I'm deadly serious am I'm going to kill you when I catch you" so the next thing your man was pure upset over this like and…

(LCIE-YW20a)

Secondly, O'Keeffe and Adolphs (2008, p. 85) state that convergence response tokens are mostly frequently found at points of convergence in conversations. They typically occur at points where:

a) participants agree;
b) participants simply converge on opinions or mundane topics, which leads them to negotiate topic boundary points collaboratively;
c) a topic can be shifted or changed.

Convergence response tokens can also be followed by a conversational closure point and thus, have a pragmatic function in that they bring about agreement and convergence, sometimes leading to topic shift. O'Keeffe and Adolphs (2008, p. 86) provide the example below (7) to

illustrate the use of the short statement "*You never know*" as a means to collaboratively round off a topic:

(7)

A: Yeah. We haven't had a night like that for a while have we?

B: No. Must have another one.

A: Silly night. [laughing] What?

B: Must have another one.

A: Well I think we will.

B: Wednesday.

A: Mm. Lifts the spirits.

B: Mm. **You never know** we might be able to get a new recruit.

A: [laughing] **You never know**.

B: [laughing]

(CANCODE-YW20b)

O'Keeffe and Adolphs (2008) conclude that convergence tokens are of a higher relational value than continuer tokens as they help maintain good relations between speakers by reinforcing commonality between them.

Thirdly, O'Keeffe and Adolphs (2008, p. 87) indicate that engagement tokens function at an affective level and signal the addressee's enthusiasm, surprise, shock and disgust at what is being said without taking over the floor. O'Keeffe and Adolphs (2008, p. 88) stress that engagement tokens are "at a much higher relational level than continuer tokens" as the former do not only sign a desire for the speaker to continue but also communicate the addressee's affective response to the speaker's message.

In conversation (8), speaker B is talking about her plans for the summer with her boyfriend in Edinburgh. The engagement token "*brilliant*" is used to convey the addressee's delight at what speaker B is saying (O'Keeffe & Adolphs, 2008, p. 87):

(8)

A: What are you going to do about the job?

B: I don't know. He says that it's going to be like Killarny and that I should get one easily enough and I've been in contact with Debenhams and they told me to send over my CV.

A: **Brilliant** Mary **brilliant**.
(LCIE-YW20a)

Furthermore, McCarthy (2003, p. 57) points out that non-minimal response tokens sometimes occur in series across speakers, "where coordinated actions produce clusters of relational signals during, for instance, (pre-)closures, and often project parallel relational convergences". The author adds that situations like pre-closures and closures enable speaker and listener roles to alternate more rapidly (McCarthy, 2003, p. 58):

(9)
B: I don't know, whatever you reckon cos I'm picking them up about gone eleven or something.
A: **Right**.
B: It's not like halfway through+
A: **Wicked**.
B: + the evening.
A: Uh huh.
B: **So** maybe I could pick you up from work and +
A: **Cool.** Yeah.
B: + go for some tea.
C: **Ace. That would be really cool**.
B: **Ah yeah**.
A: I'm defi- Yeah **that's fine for me**.
B: **Oh that's good then**.
A: **Oh excellent**.
B: I-I'm going to put that in, in pen now.
A: **Wicked**.

As for information receipt tokens, O'Keeffe and Adolphs (2008) explain that their corpora findings revealed that a small number of response tokens did not seem to fit the continuer, convergence and engagement categories, which serve relational functions. Such backchannels seemed to have more of an organisational function and were usually marked by falling pitch. In a few examples, they seemed to possess a global discourse marking function within the orientation stage of a narrative. Such response is used "as a self-imposed pragmatic marker at

which the storyteller marks a boundary where the narrative can begin now that the contingent details are clear for the participants" (O'Keeffe & Adolphs, 2008, p. 88).

In the example below (10), speaker A uses the information receipt token "*right*" at the point where she assumes all of the contingent details are in place so as to continue the story. Nonetheless, speaker B indicates that she is not ready yet by initiating a repair sequence (O'Keeffe & Adolphs, 2008, p. 88):

(10)

A: He's been in Wexford for years right. I told you he's separated didn't I? And that he has a child.

B: Yeah.

A: **Right**.

B: But he's only young isn't he?

A: He's only 29.

(LCIE- YW20a)

Following from this, O'Keeffe and Adolphs (2008, p. 89) argue that information receipt tokens are "strongly associated with asymmetrical interaction where one of the participants is a power role holder". Furthermore, some response tokens are strongly associated with particular speech events and thus, are context-specific. McCarthy (2003) indicates that "*fine*" occurs in making arrangements and reaching decisions whereas "*certainly*" occurs in replies to requests for a service or a favour, as in examples (11) and (12) respectively (O'Keeffe & Adolphs, 2008, p. 89):

(11)

A: Okay, I'll see you a bit later then.

B: **Fine**.

A: In the morning, whenever.

(CANCODE)

(12)

A: Can I have the bill, please?

B: Yes, **certainly**.

Moreover, McCarthy (2003) states that the adjectives "*excellent*", "*fine*", "*great*", "*good*", "*lovely*", "*right*" and "*perfect*" offer positive feedback to speakers and often mark topic boundaries, where speakers express their satisfaction with phases of business such as making arrangements, agreeing on courses of action and making the satisfactory exchange of information, goods and services.

Now that the forms and the functions of response tokens have been discussed, let us examine the potential candidates for response tokens in the Brazilian corpus. As adverbs had already been addressed in the previous chapters as discourse markers ("*well*", "*really*" and "*actually*") and as hedges ("*maybe*", "*probably*" and "'*jus*t"), I chose to focus on minimal forms only so as to avoid possible overlap.

I used AntConc's word list tool to view the rank positions and the raw frequencies of all the words in the Brazilian learner corpus. As the corpus is not annotated, I made a manual list of all the items that fit the form categorisation of minimal response tokens as described by O'Keeffe and Adolphs (2008). Thus, adjectives and adverbs were disregarded. I did not include the occurrences of "*yes*" and "*no*" as these are more traditionally found within answering moves to "*yes*" or "*no*" questions. Tokens with fewer than 4 occurrences were not included in the table either. Table 21 indicates the raw frequencies and the relative frequencies (per 10.000 words) of the 7 most frequent tokens that seem to fit the characterisation of minimal forms.

Table 21: Most frequent tokens fitting the description of minimal forms in the Brazilian learner corpus

Minimal forms	Raw frequencies	Relative frequency
uh	836	351.2162
yeah	123	51.6741
hum	39	16.8955
ah	28	11.7632
uhuh	18	7.5620
oh	9	3.7810
aih	4	1.6804

The above relative frequencies relate the total number of occurrences of each token, but not necessarily employed as response tokens. The most frequent minimal form used by the

Brazilian learners was "*uh*". Nonetheless, a closer look at the contexts of usage of "*uh*" corroborates the assumption that the Brazilian learners of this study used "*uh*" as hesitation device for organising one's thoughts and gaining time. "*Uh*" occurred mostly in initial and medial positions and often followed by filled and unfilled pauses. There were only 4 occurrences of "*uh*" as a one-word insert. Another form associated with hesitation was "*hum*". 18 occurrences were followed by pauses (11 unfilled/7 filled). There were only 3 occurrences of "*hum*" as a one-word insert.

The second most frequent minimal form was the short utterance "*yeah*". Out of 123 occurrences, 92 were placed in utterance initial position and 18 occurred immediately after some non-content turn-preface items or immediately after a lexical repetition. In addition, 37 occurrences in initial position displayed its use as a one-word insert. Thus, the minimal form "*yeah*" seems to have been frequently employed as a response token and then was chosen for investigation.

As for the remaining non-word vocalisations, "*ah*" was not produced as a one-word insert. However, "*ah*" occurred 9 times in utterance initial position prefacing expanded listener responses. 17 out of 18 occurrences of "*uhuh*" were placed in initial position, 12 of which were one-word inserts. Only 2 occurrences of "*oh*" displayed its use as a response token, both in initial position and followed by other response forms "*Oh, yeah!*" and "*Oh, nice!*". "*Aih*" occurred 4 times within the same speaker's turn as a marker of frustration when one cannot remember the name of something. The production of "*aih*" signals language transfer as this vocalisation is commonly used for such purpose in Brazilian Portuguese.

Considering the above analyses, I chose to narrow my investigation to the most frequent short utterance and non-word vocalisation response forms, namely "*yeah*" and "*uhuh*", respectively. The non-word vocalisations "*uh*" and "*hum*" were disregarded due to their frequent use as hesitation devices, adjacently placed to filled and unfilled pauses."*Yeah*" was the only short utterance produced by the subjects of this study as a response form, with 37 occurrences as a one-word insert while "*uhuh*" was the non-word vocalisation with the most number as one-word inserts. The contexts of usage of "*yeah*" and "*uhuh*" will be detailed in the next sub-sections so as to specify their frequency as minimal response tokens as well as the pragmatic functions performed by the same. Intonation was part of their interpretation as minimal

response forms. Although the Brazilian learner corpus is not prosodically annotated, I had listened to the interviews a number of times while transcribing the audio files. In addition, the investigation of the functions of "*yeah*"and "*uhuh*" included listening to specific extract parts again.

For comparative purposes, table 22 displays a list of possible candidates for non-minimal response tokens produced by the subjects of this study. The raw and the relative frequencies refer to the total number of occurrences in the Brazilian corpus rather than their uses as response tokens. The adverbs previously analysed in chapters 4 and 5 were not included in the table.

Table 22: Most frequent tokens fitting the description of non- minimal forms in the Brazilian learner corpus

Non-minimal forms	Raw frequencies	Relative frequency
good	147	61.7569
right	51	21.4258
okay	43	18.0649
nice	40	16.8046
interesting	21	8.8224
great	18	7.5620

As shown in table 22, the Brazilian learner corpus did not display a varied range of adjectives which can be used to express good listenership. Whether these items were actually employed as response tokens is a matter for further investigation. Nonetheless, some of these forms are among the most common non-minimal response tokens produced by native speakers of English. In the British spoken corpus CANCODE, the 10 most frequent non-minimal response tokens are as follows (McCarthy, 2002, p.59): "*right*", "*exactly*", "*fine*", "*true*", "*great*", "*definitely*", "*good*", "*lovely*", "*absolutely*" and "*gosh*". The Cambridge North American Spoken Corpus displays the following 10 most frequent items (McCarthy, 2002, p. 60): "*wow*", "*true*", "*gosh*", "*exactly*", "*absolutely*", "*right*", "*sure*", "*great*", "*definitely*".

To sum up, in this section we have examined the forms and functions of response tokens and presented the most common minimal forms produced by the subjects of this study. Minimal response tokens consist of short utterances and non-word vocalisations while non-minimal ones comprise adverbs or adjectives (functioning as pragmatic markers) and short

phrases/minimal clauses (O'Keeffe & Adolphs, 2008, p. 74). Minimal and non-minimal forms can be employed as continuer, convergence, engagement and information receipt tokens. The Brazilian learner corpus displayed the following items fitting the characterisation of minimal forms: "*uh*", "*yeah*", "*hum*", "*ah*", "*uhuh*", "*oh*" and "*aih*". After analyses of contexts of usage to confirm their uses as response tokens, the short utterance "*yeah*" and the non-word vocalisation "*uhuh*" were chosen for the empirical investigations of this chapter, to be detailed in the next sub-sections.

6.1.1. "*Yeah*"

This sub-section reports on the findings of "*yeah*" in the Brazilian learner corpus. As previously mentioned, the Brazilian data displays 123 occurrences of "*yeah*", out of which 92 are in utterance initial position and 18 occur immediately after some non-content turn-preface items or immediately after a lexical repetition. The remaining 13 occurrences were disregarded as they occur in medial position and thus, do not fit the categorisation of response tokens. Thus, I will be reporting on the uses and functions of 110 occurrences. Due to the high frequency of "*yeah*" as a listener response token in the Brazilian data, I initially present quantitative results vis-à-vis its contexts of usage and functions. Next, I illustrate specific functions and the co-occurrence of "*yeah*" with other response tokens through qualitative analyses. In the final part of this section, I round off the discussion by comparing relative frequencies in benchmark and learner corpora.

Let us initially comment on the contexts of usage of "*yeah*" in utterance initial position in the Brazilian corpus. 92 utterance initial tokens appeared in the following combinations:

a) As one-word insert sequences followed by full stops: 34 occurrences
b) As one-word insert sequences followed by exclamation marks: 3 occurrences
c) Followed by commas: 50 occurrences
d) Followed by paralinguistic elements: 5 occurrences

The above data enable us to infer that the Brazilian learners of this study used "*yeah*" mostly in utterance initial position. Having said that, out of 92 occurrences, 50 instances of "*yeah*" were produced as a preface to extended responses, possibly signalling a tendency of Brazilian

learners to extend their turns after this form. "*Yeah*" as one-word inserts was the second most frequent combination.

The distribution of the functions of 37 one-word insert sequences in utterance initial position can be summarised as follows:

a) Convergence token: 31 occurrences
b) Continuer token: 1 occurrence
c) Information receipt token: 1 occurrence
d) Other functions: 4 occurrences.

By and large, the Brazilian learners of this study produced "*yeah*" as one word inserts to perform the function of convergence, signalling agreement or simply convergence with previously mentioned ideas. In some examples, "*yeah*" as a convergence token marked topic closure or topic change. Other functions encompass "*yeah*" within an answering move following "*yes*" or "*no*" questions (3 occurrences) and "*yeah*" as a compensation strategy (1 occurrence). Thus, we can also assume that some Brazilian learners were able to adapt their discourse to the register of informal conversations by producing the informal form "*yeah*" instead of "*yes*" when replying to "*yes*" or "*no*" questions.

Let us now examine some examples of the above functions. In (1), "*yeah*" is used as a convergence response token at a point where both speakers agree:

(1)
<S14> It's horrible.
<S15> **Yeah**.
<S14> I don't like it.
(FCE P3- B1 BR-07)

In conversation (2), "*yeah*" is also used as a convergence token but only after a self-initiated repair sequence has taken place. The agreement refers back to the proposition within the first utterance "*I think the first one uh, uh will be attract a lot of people*":

(2)
<S14> Okay. All right. I think the first one uh, uh will be attract a lot of people because
<S15> <SE> overlapping</SE> The musician and music?

<S14> Yes.

<S15> **Yeah**.

(FCE P3- B1 BR-07)

In conversation (3), subjects have been discussing things people complain about in restaurants. Subject 6 initially produces the non-word vocalisation "*hum*" between two unfilled pauses marking hesitation. In his next turn, he produces "*yes*" in utterance initial position as a convergence token insert. In his third turn, he produces "*yeah*" also as a convergence token, but this time at a point leading to topic closure. The use of the exclamation mark signals a tonic prominence:

(3)

<S1> What sort of things do people complain in restaurants?

<S6> <SE> pause </SE> **Hum**, the <SE> pause </SE>, when you don't uh receive the right food that you asked.

<S7> And the food spend so much time to arrive.

<S6> **Yes**. <SE> pause </SE> Uh when you have to ask uh the waiter a lot of times the same thing.

<S7> And when are so expensive names.

<S6> **Yeah!**

<S1> Thank you. That's the end of the interview.

(FCE P4- B1 BR-03)

Conversation (4) displays 3 occurrences of "*yeah*". Subject 12 produces two consecutive convergence tokens in the form of "*yeah*", signalling stronger agreement. Subject 13 then engages in a longer turn. Next, subject 12 produces a third occurrence of "*yeah*" as a means to indicate to the speaker that he is following what is being said and to encourage the latter to continue his account. In this last occurrence, "*yeah*" is a continuer response token:

(4)

<S13> Which two? <SE> pause</SE> Maybe one free coffee it's a good idea to+

<S12> <SE> overlapping</SE> **Yeah, yeah**.

<S13> +get people to the place and uh <SE> pause</SE> good accommodations, you know, good sofas or some things like that. We're, we're not used to <SE> pause</SE> having this

kind of <SE> laughing </SE> of cafés here in Brazil. We see one of this in, on that show Friends.
<S12> **Yeah**.
<S13> Remember that? I guess it's great. And <SE> pause</SE> I don't know, I don't like too, too much about music when it's too loud in a <SE> pause</SE> in a café or in a venue or something like this.
(FCE P3- B1 BR-06)

The single occurrence of "*yeah*" as an information receipt token is shown in conversation (5). Subject 1 (the interviewer) brings the conversation to an end. Both subjects 16 and 17 use the response forms "*yeah*" and "*okay*" as information receipt tokens to acknowledge the closure of the conversation. These tokens occur within an asymmetrical interaction where the interviewer holds the power role:
(5)
<S1> Thank you. That's the end of the test.
<S16> **Yeah**.
<S17> **Okay**.
(FCE P4- B1 BR-08)

In conversation (6), we can observe the use of "*yeah*" in utterance initial position within an answering move, i.e. following a "*yes*" or "*no*" question (other function category). In the same conversation, the second occurrence of "*yeah*" prefaces an extended response within another answering move:
(6)
<S1> …Would you like to spend time in a café like this?
<S14> **Yeah**.
<S15> **Yeah,** I would love it.
<S14> I'd love too.
(FCE P4- B1 BR-07)

Lastly, also within the other function category, 1 occurrence shows "*yeah*" as a compensation strategy after the speaker has been searching for words. In the conversation below (7), subject 19 interrupts her initial turn by producing an unfilled pause that possibly marks her searching

for a specific word, namely the name of a musical instrument. Subject 18 then produces the non-word vocalisation "*huh*" twice as a continuer token encouraging speaker 19 to go on with her turn. Nonetheless, the latter does not manage to produce the word she has been searching for and, thus, produces "*yeah*" as a compensation strategy for gaining time. Subject 18 then signals his active listening by providing the lexical item "*a guitar*".

(7)

<S19> With a <SE> pause</SE> a, a good voice and a <SE> pause</SE>

<S18> **Huh, huh**

<S19> **Yeah**.

<S18> A guitar.

<S19> A guitar, yeah, yeah.

(FCE P3- B1 BR-09)

Now that the examples of "*yeah*" after full stops and exclamation marks have been presented, let us focus on *"yeah"* followed by commas. "*Yeah*" was found in utterance initial position prefacing expanded responses in 50 occurrences. The most common collocating items after the commas were as follows: "*Yeah, I…*" (10 occurrences), "*Yeah, yeah*" (8 occurrences), "*Yeah, it's…*" (3 occurrences) and "*Yeah, the…*" (3 occurrences). The distribution of the functions of "*yeah*" in utterance initial position prefacing expanded responses can be summarised as follows:

a) Convergence token: 44 occurrences
b) Other functions: 6 occurrences

The above results allow us to assume that "*yeah*" as a preface to expanded responses was also vastly used as a convergence token. These results also indicate that the Brazilian learners of this study did not produce "*yeah*" as a continuer token or an information receipt token within an extended response. On the rare occasions they produced "*yeah*" as continuer and information receipt tokens, "*yeah*" appeared as one- word inserts. The other function category encompassed 6 occurrences of "*yeah*" prefacing extended responses within an answering move following "*yes*" or "*no*" questions. This finding strengthens the assumption that some Brazilian learners are already capable of using "*yeah*" instead of "*yes*" within answering moves to mark informality. In addition, "*yeah*" seems to be viewed as a marker of the

speaker-listener world as shown by the most common collocating items "*Yes, I...*" and "*Yeah, yeah*".

Let us now examine the contexts of usage of "*yeah*" prefacing expanded responses. The conversation below (8) displays the continuation of a previous extract (conversation 6). Subjects are justifying why they would like to work in a café. Subject 14 initiates a turn with "*yeah*" as a convergence token used to simply converge with the ideas expressed in the previous turn. "*Yeah*" is then followed by the idiomatic expression *"... is not my cup of tea*", which is appropriately placed both in terms of structure and meaning:

(8)

<S1> Why?

<S15> Because I, I love café and a kind of lounging in cafés uh with musician in, in another part of the café will call my attention because uh I could uh I could stay more comfortable and read magazines and doing what I want, and after I could go to, to the other parts of the café and listen to music and it will be great.

<S14> **Yeah, coffee is not my cup of tea** but I love music so will be great to go there <SE> laughing /SE>.

(FCE P4- B1 BR-07)

In conversation (9), the use of "*yeah*" is somewhat peculiar and contradictory. Subjects are also justifying why they would like to work in a café. Subject 19 uses "*yeah*" as a convergence token to converge with the ideas presented by subject 18. However, this utterance initial "*yeah*" is then followed by a negative phrase "*I don't think so*", which refers back to the question and can be understood as "I wouldn't like to work in a café like this". This occurrence also signals the beginning of the conversation closure:

(9)

<S1> Why?

<S18>Ah because I think is a, a pleasant place uh, uh a place with, when you can know some peaceful people and serve uh, serve uh drinks to people.

<S19> **Yeah, I don't think so**. I think when you work in this place it's a little uh complicated because you have to, to uh <SE> pause</SE> to know the people, you are a, a, a employee and the, the, the people that are you serving are uh <SE> pause</SE> not all the people are in peaceful, sometimes they are very <SE> chuckles</SE> , I think it's, it's hard

to, to work with uh, with restaurants or, or eat or this kind of places. I think you have a lot of patience and I am not so patient <SE> chuckles</SE>, yeah.
<S1> Thank you. That's the end of the test.
<S19> You are welcome.
(FCE P4- B1 BR-09)

A similar contradiction is also perceived in the occurrences of "*yeah*" followed by other attitudinal verbs in negative forms such as "*I don't like*" and "*I don't know*" or followed by "*but*" as in the conversation below. In (10), subjects have been discussing why young people usually go to different places to relax than older people. Subject 15 implies that preferences are age-dependent. Subject 14 then produces "*yeah*" as a convergence token to mark basic convergence with this idea but then disagrees with it by saying that preferences have more to do with individual personalities:
(10)
<S15> And while my, my grandparents love to stay home and <SE> laughing /SE> for them it's great to, to watch TV together so it's totally different for teenagers.
<S14> **Yeah, but** for my, for me and my mom we go to spend time in the <SE> laughing /SE> in the mall <SE> chuckles /SE> so it's not so different with age , even with the personalities, I guess.
<S1> Thank you. That's the end of the test.
(FCE P4- B1 BR-07)

In conversation (11), there are 3 occurrences of "*yeah*". Subjects are discussing the importance of spending time and money on free time activities. Subject 12 produces two consecutive occurrences of "*yeah*" as a convergence token indicating stronger agreement with the idea of spending money on travelling. Subject 13 then produces "*yeah*" combined with the non-minimal form "*for sure*" as a means to signal strong agreement. This last occurrence of "*yeah*" marks the closure of the topic, bringing the conversation to its end:
(11)
<S12> …I think it's a <SE> pause</SE> , I think you should travel , you should spend your money to <SE> pause</SE>+
<S13> <SE> overlapping</SE> know another country
<S12> <SE> overlapping</SE>**Yeah, yeah**.

<S13> +to, to visit some different places, to , ah if you should like to go sports, I don't know.
<S12> To have contact with different cultures.
<S13> **Yeah, for sure**.
<S1> Okay, thank you. That's the end of the test.
<S13> Thank you very much .
<S12> Thank you.
(FCE P4- B1 BR-06)

Within the other function category of "*yeah*" in utterance initial position preceding an expanded response, all 6 occurrences displayed "*yeah*" within an answering move following "*yes*" or "*no*" questions, as in conversation (12):

(12)
<S1> … have you got plans for this weekend?
<S16> For this weekend? **Yeah**, I will go home <SE> laughing /SE> and on Sunday but on Saturday I will go out with my friends because it's the last day I will be here, I will return after the holidays <SE> laughing /SE>but this is my plan for this weekend <SE> laughing /SE>.
(FCE P1- B1 BR-08)

Lastly, the occurrences of "*yeah*" followed by paralinguistic elements included 3 occurrences followed by laughter and 2 occurrences followed by unfilled pauses. In all 5 occurrences, "*yeah*" was also employed as a convergence token. In conversation (13), "*yeah*" is produced in overlapping talk and followed by laughter. Both paralinguistic phenomena seem to strengthen the level of high convergence:

(13)
<S4> I think is a good idea. You can have a live music.
< S5> The free coffee as well <SE> laughing </SE>.
<S4> <SE> overlapping</SE> **Yeah** <SE> **laughing** </SE>.
(FCE P3- B1 BR-02)

As for the occurrences of "*yeah*" immediately after some non-content turn-preface items or immediately after a lexical repetition, these appeared in the following collocations:

a) Immediately following an initial "*yeah*" : 10 occurrences

b) Immediately after a non-content preface: 2 occurrences
c) Immediately after the repetition of a lexical item from the previous speaker's turn: 6 occurrences.

With the exception of one single occurrence of "*yeah*" performing the continuer function, the remaining 17 occurrences displayed "*yeah*" as a convergence token. These results lead us to believe that the Brazilian learners of this study were able to produce "*yeah*" at relevant points which strengthened active listenership. In addition, the occurrences of "*yeah*" were largely employed as a means to agree or convergence with previously presented ideas.

Let us now examine some examples of "*yeah*" immediately following prefaces. In (14), the first turn of speaker 4 seems to point to language transfer through the use of "*é*", a Brazilian vocalisation frequently used when searching for words or when organising ideas. In the next turn, subject 5 produces "*yeah*" preceded by another minimal form, namely the non-word vocalisation "*oh*". The co-occurrence of the minimal forms "*oh*" and "*yeah*" seems to indicate a much stronger level of agreement:

(14)

<S4> … Uh, in this <SE> pause</SE>have international menus **é**, for example, here in Brazil, in these times of our World Cup can be a good <SE> pause</SE>

<S5> **Oh yeah**!

< S4> a good uh <SE> pause</SE> tools, tool to uh call uh visitors and uh customers to go to the restaurant to a, if you wanna some problems to identify some specific foods you can find in your language, language, so can be a good attractive for the, for the business and restaurant.

(FCE P3- B1 BR-02)

Conversation (15) displays the single occurrence of "*yeah*" as a continuer token preceded by a lexical repetition. Subject 13 engages in a narrative and produces "*Starbucks*" as its last element. Subject 12 then repeats the proper noun and produces "*yeah*" indicating that he is following the story and wants the speaker to continue it:

(15)

<S13> + and things like that cause I am addicted to <SE> pause</SE> coffee and coffee shops. Uh I get addicted to that when I went to New York and I<SE> pause</SE> and I went for the first time in a, in a **Starbucks**.
<S12> **Starbucks**, **yeah**.
(FCE P4- B1 BR-06)

Now that we have discussed all the different contexts of usage of "*yeah*", let us summarise all results. Table 23 indicates the raw frequencies and the distribution of "*yeah*" among functions in the Brazilian corpus.

Table 23: Distribution of "*yeah*" among functions in the Brazilian learner corpus

Yeah	Utterance initial	Immediately after a preface	Percentage
Convergence tokens 97 occurrences	80	17	88.19%
Within an answering move 9 occurrences	9	X	8.19%
Continuer tokens 2 occurrences	1	1	1.82%
Information receipt tokens 1 occurrence	1	X	0.90%
As a compensation strategy 1 occurrence	1	X	0.90%
Total: 110 occurrences	92	18	100%

Overall, it is possible to conclude that the Brazilian subjects of this study employed "*yeah*" largely as a convergence token (88.19%), signalling a tendency for agreement within preference structure. Nonetheless, some occurrences of "*yeah*" as a convergence token displayed a contradiction, as in the cases of "*yeah*" followed by attitudinal verbs in negative forms or by "*but*". In those examples, the token "*yeah*" seems to have been employed as an informal politeness formula to reinforce common ground despite disagreement[51]. As for the

[51] Within positive politeness, Brown and Levinson (1987, p. 113) describe strategy number 6 as a means to avoid disagreements, indicating the existence of token agreement and pseudo-agreement.

other functions, a number of occurrences (8.19%) showed a similar use to "*yes*", within answering moves to "*yes*" or "*no*" questions. Conversely, there were occasional occurrences of "*yeah*" as a continuer token (1.82%) and as an information receipt token (0.90%).

Now let us compare the frequencies of "*yeah*" with native speaker data. Table 24 displays the raw and the relative frequencies of "*yeah*" in utterance initial and in other positions (medial and utterance final) in the two benchmark corpora and in the Brazilian learner corpus. All occurrences of "*yeah*" have been included irrespective of specific functions. Relative frequencies have been normalised to 10,000 words. The CQP web server allows users to conduct case-sensitive searches. Thus, the utterance initial native speaker figures refer to "*yeah*" in upper case while the other position figures relate to its occurrences in lower case. Due to the way the face-to-face conversation from the 1990s sub-corpus had been originally designed, I was unable to distinguish between medial and utterance final positions. Thus, both positions have been grouped together as shown in the benchmark data.

Table 24: Raw and relative frequencies of "*yeah*" in 3 corpora

	BNC		F-2-F		Brazilian corpus	
Yeah	Raw frequencies	Relative frequency	Raw frequencies	Relative frequency	Raw frequencies	Relative frequency
Utterance-initial	**6,586**	**57.7691**	**1,652**	**60.8785**	**110**	**46.2126**
Other positions	2,214	19.4201	389	14.3352	13	5.4614
Total	8,800	77.1892	2,041	75.2137	123	51.6774

In the three corpora, one can notice a far higher frequency of "*yeah*" in utterance initial position, which corroborates the assumption that "*yeah*" is mostly used as a response token. In the specialised sub-corpus, its frequency in utterance initial position is even higher, indicating a very strong association with face-to-face conversations. Despite producing "*yeah*" in utterance initial position in a lower frequency if compared to those in native speaker data, the Brazilian learners of this study showed a consistent use of "*yeah*" as a listener response token in this position. Thus, one can conclude that Brazilian learners of this study produced "*yeah*" less frequently than native speakers did but they displayed a consistent use of "*yeah*" as a listener response token.

The Spanish component of LINDSEI displayed 1,166 occurrences of "*yeah*" (relative frequency 103.2580). The relative frequency was considerably higher than those of native speakers and much higher than that of the Brazilian learners. As all instances of "*yeah*" appeared in lower case in the Spanish corpus, a manual analysis of 1,166 occurrences would be required so as to specify the number of examples in utterance initial position. As other learner data is occasionally referred to for illustrative purposes only, LINDSEI data was not included in table 24.

Now let us finalise this discussion by presenting the most frequent collocations of "*yeah*" as produced by native speakers. The collocating items in the Brazilian corpus have already been presented through the qualitative analyses. Table 25 indicates the 12 most frequent collocating items with "*yeah*" in utterance-initial position in both benchmark corpora. The percentage figures refer to the collocation distribution among other collocations.

Table 25: Most frequent collocations with "*yeah*" in utterance-initial position in the benchmark corpora

BNC			**F-2-F**		
Collocation	Raw frequencies	Percentage	Collocation	Raw frequencies	Percentage
Yeah.	3562	54.08%	Yeah I	199	12.05%
Yeah,	1020	15.49%	Yeah < (punct.)	185	11.20%
Yeah I	256	3.89%	Yeah Yeah	161	9.75%
Yeah but	164	2.49%	Yeah It	74	4.48%
Yeah!	138	2.10%	Yeah but	72	4.36%
Yeah?	132	2.00%	Yeah that	67	4.06%
Yeah and	115	1.75%	Yeah Well	57	3.45%
Yeah it	83	1.26%	Yeah And	45	2.72%
Yeah that	79	1.20%	Yeah yes	41	2.48%
Yeah well	78	1.18%	Yeah mm	40	2.42%
Yeah yeah	74	1.12%	Yeah Uhm	39	3.36%
Yeah you	69	1.05%	Yeah You	36	2.18%

Despite differences in design of both corpora and in the rank position of collocations, one can notice that the collocating items are practically the same. Nonetheless, "*yeah*" is clearly and

mostly used as a single word response form in the BNC spoken sub-corpus whereas the specialised face-to-face conversation sub-corpus indicates a higher percentage of "*yeah*" followed by "*I*", possibly signalling a tendency of use of "*yeah*" preceding attitudinal verbs. Furthermore, two non-word response forms "*mm*" and "*uhm*"are also among the most frequent collocations in the specialised sub-corpus. In both native speaker corpora, "*yeah*" followed by "*but*" is also among the most frequent collocations, reinforcing its use as a pseudo-agreement response token. By comparing collocating items in the native speaker corpora with the Brazilian data, it is possible to state that the subjects of this study produced most collocating items, with the exception of "*Yeah?*", "*Yeah that*", "*Yeah well*", "*Yeah mm*" and " *Yeah uhm*". This comparison suggests that the Brazilian learners of this study employed "*yeah*" as a response token within similar contexts of usage to those of native speakers.

To conclude, in this section we have investigated the functions of 110 occurrences of "*yeah*" in the Brazilian learner corpus and compared frequencies with native speaker data. The Brazilian subjects of this study employed "*yeah*" largely as a convergence token, signalling a tendency for agreement within preference structure. Nonetheless, some occurrences of "*yeah*" as a convergence token displayed a contradiction as in the cases of "*yeah*" followed by attitudinal verbs in negative forms or by "*but*". A number of occurrences showed a similar use to "*yes*" within answering moves to "*yes*" or "*no*" questions. There were occasional occurrences of "*yeah*" as a continuer token and as an information receipt token. Comparisons with native speaker data indicated that the subjects of this study produced "*yeah*" less frequently than native speakers did but displayed a consistent use of "*yeah*" as a listener response token, in similar contexts of usage.

6.1.2. "*Uhuh*"

This sub-section reports on the findings of the non-word "*uhuh*" in the Brazilian learner corpus (also spelt as "*uh-huh*" and "*uhu"* in other corpora). As previously mentioned, the decision for investigating "*uhuh*" as a minimal response token was based on its highest raw frequency as a one-word insert, if compared to the frequencies of the other non-word vocalisations produced by the subjects of this study. The Brazilian corpus displays 18 occurrences of "*uhuh*", out of which 17 are in utterance initial position and 1 is in medial

position. Therefore, I will be focusing on the contexts of usage and on the functions of the occurrences in utterance initial position.

Table 26 indicates the contexts of usage *of "uhuh"* in utterance initial position, namely as a one-word insert sequence followed by full stops, followed by commas preceding extended responses and preceding paralinguistic elements. It also displays the distribution of "*uhuh*" among functions in specifc contexts.

Table 26: Distribution of "*uhuh*" among functions in utterance initial position in the Brazilian learner corpus

Uhuh **Utterance initial**	One- word inserts	Followed by commas	Followed by chuckles	Percentage
Convergence tokens 8 occurrences	6	1	1	47.05%
Continuer tokens 6 occurrences	6			35.30%
Information receipt tokens 3 occurrences	2	1		17.65%
Total: 17 occurrences	14	2	1	100%

The above results attest a much higher frequency of "*uhuh*" as a single word response form. There were only 2 occurrences of "*uhuh*" preceding extended responses and 1 occasional occurrence preceding paralinguistic elements. Thus, one can conclude that the subjects of this study employed "*uhuh*" in utterance initial position largely as one-word inserts (82.35%). Furthermore, the qualitative analyses of "*uhuh*" indicated 3 response functions, namely as a convergence token, as a continuer token and as an information receipt token. While the Brazilian learners used "*uhuh*" more frequently as a convergence token, its context of usage varied. The second most common function was "*uhuh*" as a continuer token, entirely as one-word insert sequences. The least frequent use was as an information receipt token, both as a one word insert sequence or preceding an extended response.

Let us now exemplify some of the above functions. In the conversation below (1), subjects have just received the instructions for a speaking task. The first turns of both subjects 15 and 14 illustrate the uses of the non-minimal forms "*okay*" and "*all right*" as information receipt tokens. Subject 15 then self- initiates a repair sequence, which is subsequently followed by a

series of convergence tokens: "*yeah*", "*uhuh*" and "*uhuh*". This series of convergence tokens indicates a high level of agreement with the ideas proposed by subject 14:

(1)

<S1> Thank you. Now I'd like you to talk about something together for about three minutes. I'd like you to image that a local café wants to attract more people. Here are some suggestions they are considering. <SE> handing out visual prompts </SE>. First talk to each other about how successful these suggestions might be. Then decide which two would attract most people. All right?

<S15> **All right**.

<S14> **Okay**. **All right**. I think the first one uh, uh will be attract a lot of people because

<S15> <SE> overlapping</SE> The musician and music?

<S14> Yes.

<S15> **Yeah**.

<S14> And everybody likes music so <SE> pause</SE>is important to go to a place uh that have a <SE> pause</SE>a group playing alive so.

<S15> **Uhuh**.

<S14> I think it's great.

<S15> **Uhuh**. Every week, weekend I,…

(FCE P3- B1 BR-07)

Similarly to the previous extract, in conversation (2) the interviewer delivers the rubrics to the task that follows and self-imposes an information receipt token to confirm understanding "*All right*". Subjects 19 and 18 acknowledge receipt through the response tokens "*all right*" and "*right*" respectively. Subject 19 then engages in a longer turn, which is then followed by subject's 18 production of "*uhuh*" as a convergence token. "*Uhuh*" also marks topic closure:

(2)

<S1> Thank you. Now I'd like you to talk about something together for about three minutes. I'd like you to image that a local café wants to attract more people. Here are some suggestions they are considering. <SE> handing out visual prompts </SE>. First talk to each other about how successful these suggestions might be. Then decide which two would attract most people. <SE> pause</SE> **All right**?

<S19> **All right**.

<S18> **Right**.

<S19> I think uh <SE> pause</SE> it's important the <SE> pause </SE> place uh in this case I think when you have a open place with uh, with a good uh, uh location and you have some trees and some, some things that uh that uh help the, the, the place are more comfortable.
<S18> **Uhuh**.
<S19> And what do you think about music?
(FCE P3- B1 BR-09)

Conversation (3) displays one occurrence of "*uhuh*" as a convergence token surrounded by laughter. The co-occurrence of "*uhuh*" with this paralinguistic phenomenon seems to strengthen the degree of agreement and to reinforce common ground:
(3)
<S18> Uh <SE> pause</SE> if you want to, to, to attract much, much, much people <SE> laughing </SE>in your coffee maybe you can put the television with the uh football game.
<S19> <SE> laughing </SE> Coffee with a video game.
<S18> <SE> chuckles</SE>. But probably the, the, the café will uh will <SE> pause</SE> will lost some characters of a coffee.
<S19> **Uhuh** <SE> chuckles</SE>. And I think the, the, the cost, the price, it's not the, the important thing…
(FCE P3- B1 BR-09)

Conversation (4) is the continuation of the previous extract (conversation 3). The minimal forms "*uhuh*" and "*yeah*" occur as triplets, performing the function of convergence. The production of three consecutive tokens within overlapping talk also indicates a higher level of agreement:
(4)
<S18> Uh maybe a, maybe another possibility is a, to transform the, to change the café in, on a night café like this picture here.
<S19> **Uhuh**.
<S18> And change all the, the layout of café <SE> pause</SE> uh to, to, to attract more and more people, I don't know the <SE> pause</SE>, I don't know the <SE> pause</SE>
<S19> To <SE> pause</SE>
<S18> Uh the, the <SE> pause</SE>

<S19> To the place?
<S18> Yeah, to the place, the, the, the size of the place <SE> laughing </SE> +
<S19> <SE> overlapping</SE> **uhuh, yeah, yeah**
(FCE P3- B1 BR-09)

In conversation (5), the interviewer is delivering the rubrics to the task. Subject 4 produces "*uhuh*" as a continuer token, signalling that he is listening attentively to the instructions and encouraging the interviewer to continue her explanation. She does so and subject 4 produces a second "*uhuh*" when the former ends her turn. In this last occurrence, "*uhuh*" performs the function of information receipt and can be understood as: "Okay, I know now what I am supposed to do":

(5)
<S1> … In this part of the test I'm going to give you each of you two photographs. I'd like you to talk about your photographs on your own for about a minute and also to answer a short question about your partner's photograph. …, it's your turn first, here are your photographs.
<S4> **Uhuh.**
<S1> They show people making music in different ways. <SE> pause </SE> I'd like you to compare the photographs and say why you think music is important to the different groups of people.
<S4> **Uhuh**. So in the first one….
(FCE P2- B1 BR-02)

In the conversation below (6), subject 10 is describing New Yorkers' way of life. Subject 11 produces "*uhuh*" in overlapping talk to mark her interest in the speaker's account and to encourage him to continue it. In her next turn, she produces the non-minimal form "*nice*" as an engagement token to express her delight at what subject 10 has just said:

(6)
<S10> + uh an expresso. And I <SE> pause</SE> but I also like to order a café and go outside, in the winter.
<S11> Really?
<S10> Really, really. It's like a New York way of life. Go to the subway with a café and uh maybe with a kindle +
<S11> **Uhuh.**

<S10> + in the <SE> pause</SE> other hand.
<S11> **Nice**.
<S10> Really something.
(FCE P4- B1 BR-05)

Now that some of the contexts of usage of "*uhuh*" have been presented, let us turn our attention to native speaker data. Unfortunately, there were very few occurrences of "*uhuh*" displayed in both benchmark corpora. In the BNC Sampler, 2 occurrences of "*uhuh*" spelt as "*uh-huh*" were found in its written component. Both examples consisted of one-word inserts in narratives. There was not a single occurrence of "*uhuh*" in its spoken component. Likewise, the Diachronic Corpus of Present-Day Spoken English presented 10 occurrences of "*uhuh*", mostly in medial position. There was not a single occurrence in its sub-corpus "face-to-face conversations from the 1990s". A further investigation of "*uhuh*" was then carried out in the Corpus of Contemporary American English (COCA). As previously mentioned in section 3.3.3, COCA is a 450 million-word corpus that includes both written and spoken genres. Again, there were only 5 occurrences of "*uhuh*" as one-word inserts in narratives, as shown in Figure 7:

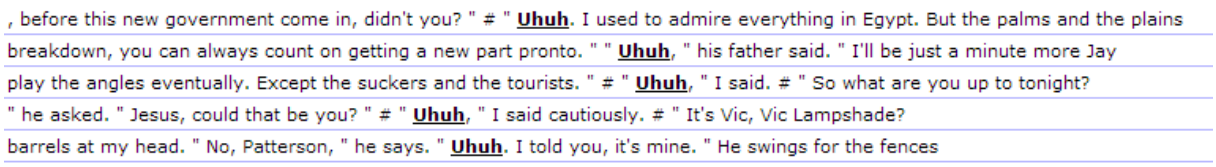


Figure 7: Extract of concordance lines for "*uhuh*" in COCA

Insufficient corpus-based evidence of the non-word "*uhuh*" may be due to different reasons. Firstly, the transcription of non-word vocalisations lends itself to multiple forms. Thus, a non-word form like "*uhuh*" may be transcribed in different ways. Secondly, non-word vocalisations may not necessarily be salient to people transcribing spoken corpora, and thus, they may go unnoticed. Thirdly, vocalisations may not be the focus of interest of researchers building spoken corpora and consequently, they may be simply ignored.

For comparative purposes, the Spanish component of LINDSEI displayed 93 occurrences of "*uhu*" (8.2358 per 10,000 words), which is a variant of "*uhuh*". Out of the 93 occurrences, 80 were clearly produced in utterance initial position while other occurrences were also found in utterance initial position but preceded by other response tokens. The Brazilian learner corpus

displayed a relative frequency of 7.5620 (per 10,000 words), considering all 18 occurrences. Despite insufficient native speaker data, one can suggest that the Brazilian learners of this study produced "*uhuh*" less frequently than the Spanish learners did. However, they were able to employ it as a listener response token performing the convergence, continuer and information receipt functions at relevant places in conversations.

In this section, we have investigated the contexts of usage and the distribution among functions of "*uhuh*" as a response token. The subjects of this study employed "*uhuh*" in utterance initial position largely as one-word inserts (82.35%). There were 2 occurrences of "*uhuh*" preceding extended responses and 1 occurrence preceding paralinguistic elements. The Brazilian learners used "*uhuh*" more frequently as a convergence token. The second most common function was "*uhuh*" as a continuer token. The least frequent function was as an information receipt token. It was concluded that the Brazilian learners employed "*uhuh*" as a listener response token at relevant places in conversations.

6.2 DISCUSSION AND RECOMMENDATIONS

So far we have presented the contexts of usage and response functions of the minimal forms "*yeah*"and "*uhuh*" in utterance initial position in the Brazilian learner corpus. Let us now summarise these findings in the light of the specific research sub-question addressed in this chapter.

RSQ3: *What are the most common minimal response tokens used to express good listenership by Brazilian learners at CEFR B1? How do their frequency of use and pragmatic functions compare to those of native speakers?*

Let us initially revise the minimal forms present in the Brazilian data. As shown in table 21, the Brazilian learner corpus displayed the following most frequent tokens fitting the description of minimal forms (O'Keeffe and Adolphs , 2008): "*uh*" (351.2162), "*yeah*" (51.6741), "*hum*" (16.8955), "*ah*" (11.7632), "*uhuh*" (7.5620), "*oh*" (3.781) and "*aih*" (1.6804). Nonetheless, relative frequencies (per 10,000 words) relate to the total number of occurrences of each token; not all occurrences of individual forms were employed as response tokens, as previously discussed. A closer look at the contexts of usage of individual forms

revealed that "*yeah*" was the only short utterance produced by the subjects of this study as a response token, with 37 occurrences as one-word inserts. "*Uhuh*" was the non-word vocalisation most frequently used as a response token: 17 out of 18 occurrences of "*uhuh*" were placed in initial position, 12 of which were one-word inserts. Therefore, the minimal forms "*yeah*" and "*uhuh*" were chosen for the empirical investigations.

Let us now compare the distribution of the minimal forms under investigation in utterance initial position as produced by individual learners. Table 27 displays "*yeah*" and "*uhuh*" raw frequencies per subject.

Table 27: Distribution of minimal response tokens per subject

Brazilian learners	**Yeah**	**Uhuh**
Subject 2		
Subject 3		
Subject 4	4	4
Subject 5	5	
Subject 6	1	
Subject 7	2	
Subject 8		
Subject 9	4	
Subject 10	6	
Subject 11	3	2
Subject 12	9	2
Subject 13	5	
Subject 14	6	1
Subject 15	12	2
Subject 16	18	1
Subject 17	8	
Subject 18	6	2
Subject 19	12	3
Subject 20		
Subject 21	9	
Total	**110**	**17**

The above results allow us to conclude that 80% of the subjects of this study produced the short utterance "*yeah*". 40% of the Brazilian learners employed both minimal forms "*yeah*" and "*uhuh*" to express good listenership. Thus, one can conclude that non-word vocalisation

"*uhuh*" was never produced in isolation as a response token by these subjects, who also employed "*yeah*" with much higher frequencies, with the exception of subject 4, who produced both forms with the same raw frequency.

Let us now discuss the findings of specific forms. This study investigated 110 occurrences of "*yeah*", out of which 92 were in utterance initial position and 18 occurred immediately after some non-content turn-preface items or immediately after a lexical repetition, also in utterance initial position. Native speaker data attested a far higher frequency of "*yeah*" in utterance initial position than in medial and final positions. It was concluded that the subjects of this study produced "*yeah*" less frequently than native speakers did but they displayed a consistent use of "*yeah*" as a listener response token in utterance initial position. Relative frequencies (per 10,000 words) were as follows:

a) BNC Sampler spoken sub-corpus: utterance initial position (57.7691), other positions (19.4201)

b) Face-to-face conversations from the 1990s: utterance initial position (60.8785), other positions (14.3352)

c) Brazilian learner corpus: utterance initial position (46.2126), other positions (5.4614)

The functions of "*yeah*" as a listener response token in utterance initial position in the Brazilian data can be summarised as follows:

a) Convergence token: 88.19%

b) Continuer token: 1.82%

c) Information receipt token: 0.90%

Other uses of "*yeah*" in utterance initial position included the following functions:

a) Similar to "*yes*" within answering moves ("*yes*" or "*no*" questions): 8.19%

b) As a compensation strategy: 0.90%

Considering the above figures, one can conclude that the Brazilian learners of this study produced "*yeah*"in utterance initial position largely as a response token (90.91%). In addition, 88.19% of the occurrences in initial position displayed its use as a convergence token, signalling a tendency for agreement within preference structure. Some occurrences of "*yeah*" as a convergence token indicated its use as an informal politeness formula employed to

reinforce common ground despite disagreement. There were occasional occurrences of "*yeah*" as a continuer token (0.90%) and as an information receipt token (0.90%).

As for the second item under investigation, this study analysed 17 occurrences of "*uhuh*" in utterance initial position present in the Brazilian learner corpus:

a) As one-word insert sequences followed by full stops: 14 occurrences
b) Followed by commas: 2 occurrences
c) Followed by paralinguistic elements: 1 occurrence

The distribution of "*uhuh*" among functions in utterance initial position in the Brazilian learner corpus was as follows:

a) Convergence token: 47.05%
b) Continuer token: 35.30%
c) Information receipt token: 17.65%

The above results allow us to conclude that all the occurrences of "*uhuh*" in utterance initial position showed its use as a listener response token. The subjects of this study employed "*uhuh*" more frequently as a convergence token. Nonetheless, its use as a continuer token was also evident. There were occasional occurrences of "*uhuh*"as an information receipt token.

Due to insufficient native speaker corpus evidence of "*uhuh*" (as discussed in section 6.1.2), its relative frequency was compared to that of more advanced learners (Spanish component of LINDSEI). It was concluded that the Brazilian learners of this study produced "*uhuh*" less frequently than the Spanish learners did. However, they were able to employ it as a listener response token performing the convergence, continuer and information receipt functions at relevant places in conversations. Relative frequencies (per 10,000 words) were as follows:

a) Brazilian learner corpus: 7.5620
b) Spanish component of LINDSEI: 8.2358

Lastly, let us compare the frequencies of the minimal forms "*yeah*" and "*uhuh*" vis-à-vis their functions as response tokens. Table 28 indicates the raw frequencies and the relative frequencies (per 10,000 words) of both minimal forms in the Brazilian learner corpus per function.

Table 28: "*Yeah*" and "*uhuh*" frequencies as response tokens in utterance initial position in the Brazilian learner corpus

	Yeah		**Uhuh**	
Response tokens functions	Raw frequencies	Relative frequency	Raw frequencies	Relative frequency
Convergence tokens	97	40.7511	8	3.3609
Continuer tokens	2	0.8402	6	2.5206
Information receipt tokens	1	0.4201	3	1.2603
Total	**100**	**42.0115**	**17**	**7.1419**

The above data allow us to infer that the subjects of this study employed both "*yeah*" and "*uhuh*" more frequently as convergence tokens. However, they used the short utterance "*yeah*" twelve times more frequently than the non-word vocalisation "*uhuh*" in this function. Conversely, the Brazilian learners of this study used the non-word vocalisation more frequently as a continuer token and as an information receipt token. As previously mentioned in section 6.1, convergence tokens are of a higher relational value than continuer tokens as they help maintain good relations between speakers by reinforcing commonality between them (O'Keeffe and Adolphs , 2008). Thus, one may assume that the Brazilian learners of this study displayed a preference for the short utterance form to express convergence. As for speech management functions, namely continuer and information receipt, learners seemed to prefer the non-word vocalisation.

Data from young British women (O'Keeffe and Adolphs, 2008, p.90) indicated that convergence tokens occurred twice as frequently as engagement tokens, ten times as frequently as continuer tokens and 11 times as frequently as information receipt tokens. Adding the occurrences of both minimal forms, the Brazilian data displayed even larger differences: convergence tokens occurred 13 times as frequently as continuer tokens and 26 times as frequently as information receipt tokens. Relative frequencies were as follows (per 10,000 words):

a) Convergence: 44.1120

b) Continuer: 3.3609

c) Information receipt: 1.6804

These results attest that the Brazilian learners of this study employed the minimal forms largely as a means to express agreement or convergence. Continuer and information receipt occurrences were much less frequent among functions. In addition, the minimal forms were not employed to express engagement.Considering that engagement tokens express genuine emotional responses such as surprise, shock, horror, sympathy and empathy, speakers tend to employ non-mininal forms (adjective and adverbs), short statements and repetitions to convey engagement. Thus, a further investigation was then manually carried out in order to identify forms of engagement tokens in the Brazilian corpus. Within the 10 interviews, there were only 8 occurrences of engagement tokens (3.3609 per 10,000 words): "*Exactly*" (2 occurrences), "*Very nice*" (1 occurrence), "*Oh nice*" (1 occurrence), "*Nice*" (1 occurrence), '*It's horrible*" (1 occurrence), "*Great*" (1 occurrence), "*Wow*" (1 occurrence). It is possible to infer that the subjects of this study produced engagement 13 times less frequently than convergence tokens and with the same frequency as continuer tokens, in comparison with "*yeah*"and "*uhuh*" frequencies. All in all, results indicate possible underuse of engagement, continuer and information response tokens by the Brazilian learners of this study. However, these findings are to be considered partial since a comprehensive account of response tokens and functions would require the analyses of all the minimal and non-minimal forms in the Brazilian corpus.

The above discussion leads us to conclude that the subjects of this study would benefit from classroom activities that offer learners the opportunity to employ more varied minimal and non- minimal response tokens in general. The analyses of the forms fitting the description of minimal and non -minimal forms showed a limited range, as discussed in section 6.1. In addition, the subjects of this study did not produce engagement, continuer and information receipt tokens in the same proportion among functions as native speakers did. Thus, learners would also gain from activities aiming at different forms to emotionally respond to previous communicative messages, to encourage speakers to continue their ongoing turns and to acknowledge understading.

Considerinng Thornbury and Slade's (2006, p. 296) three part cycle "exposure, instruction and practice", a similar approach to that of discourse markers (as suggested in section 4.2)

could be implemented. Teachers can previously select an authentic conversation extract that includes a range of response tokens and that is at a level of listening proficiency their learners can cope with. Next, teachers can create a gapped version of the extract, deleting the response tokens. In class, teachers can initially play the conversation for the understanding of the situation, asking students questions focusing on general comprehension (i.e. "gist"), for instance:

1. What is the relationship between the speakers?
2. What are the speakers talking about?
3. Where does the conversation take place?

Once situational meaning has been established, teachers can give out the gapped version of the transcript and play the extract again for students to complete the gaps with specific tokens. Following this, teachers can ask students what those forms have in common and then present their functions in lay terms: e.g. expressing agreement, expressing emotions and interest, encouraging speakers to continue talking and confirming understanding. After that, teachers can brainstorm other forms which can also be used for these purposes. Native speaker data (CANCODE) indicates the following most common one-word response tokens (McCarthy, 2002, p.59), all bearing positive connotations: "*right*", "*exactly*", "*fine*", "*true*", "*great*", "*definitely*", "*good*", "*lovely*", "*absolutely*", "*gosh*", "*wow*", "*really*", "*sure*", "*cool*", "*brilliant*", "*excellent*", "*wonderful*", "*certainly*", "*marvelous*", "*perfect*" and "*quite*". As a practice activity, learners can then roleplay a similar situation, employing response tokens so as to express good listenership.

An alternative practice activity would be for teachers to narrate an interesting true story that happened to them some time ago, but telling the story line by line. Each time teachers finish individual lines, students need to react to the information presented by expressing interest or by asking follow-up questions. Once teachers finish their story, learners can then perform the same activity in pairs. This activity in particular would offer learners the opportunity to employ continuer and engagement tokens.

<u>For instance:</u>

Teacher: "I saw the most incredible film last weekend".

Student A: "Really?"

Student B: "What was that?"

Teacher: ….

To conclude, in this section we have summarised corpus-based findings vis-à-vis the relative frequencies and the response functions of the minimal forms "*yeah*" and "*uhuh*" in utterance initial position in the Brazilian learner corpus and compared results with native speaker data. We have discussed findings in the light of the specific research sub-question addressed in this chapter. We have also proposed some pedagogical recommendations.

SUMMARY

This chapter has examined the most common minimal response tokens used to express good listenership by Brazilian learners at CEFR B1. Good listenership refers to a set of words that routinely perform interactional/relational responsive functions (McCarthy, 2002). Minimal response tokens consist of short utterances and non-word vocalisations (O'Keeffe & Adolphs, 2008, p. 74). In addition, they are minimal forms as they have no homonyms in other word classes. Non-minimal response tokens are adverbs or adjectives functioning as pragmatic markers and short phrases/minimal clauses (O'Keeffe & Adolphs, 2008, p. 74). Considering that adverbs had already been addressed in the previous chapters of this thesis, only minimal forms were chosen as the focus of investigation of chapter 6 so as to avoid possible overlap.

Response forms can be categorised as continuer, convergence, engagement and information receipt tokens (O'Keeffe and Adolphs, 2008). Continuer response tokens are facilitative as they maintain the flow of talk by encouraging the speaker to continue. Convergence response tokens are mostly frequently found at points where participants agree or simply converge on opinions or mundane topics or where a topic can be shifted or changed. Engagement tokens function at an affective level and signal the addressee's enthusiasm, surprise, shock and disgust at what is being said without taking over the floor. Information receipt tokens have more of an organisational function and are usually marked by falling pitch.

AntConc's word list tool was used to view the rank positions and raw frequencies of all the words in the Brazilian learner corpus. As the corpus is not annotated, a manual list was made of all the items that fit the form categorisation of minimal response tokens as described by O'Keeffe and Adolphs (2008). The most frequent minimal form used by Brazilian learners

was "*uh*". Nonetheless, a closer look at the contexts of usage of "*uh*" indicated its use as a hesitation device, similarity to "*hum*", also present in the Brazilian production. The second most frequent minimal form was "*yeah*". Analyses of its context of usage indicated its function as a response token. "*Ah*" was not produced as a one-word insert. "*Uhuh*" was largely produced as a reponse token. Only 2 occurrences of "*oh*" displayed its use as a response token. "*Aih*" was only produced within the same speaker's turn. Following from these analyses, "*yeah*" and "*uhuh*" were chosen for the empirical investigations as the most frequent short utterance and non-word vocalisation response forms, respectively.

The sub-sections of chapter 6 investigated the contexts of usage and functions of "*yeah*" and "*uhuh*" in the Brazilian learner corpus. The corpus displayed 110 occurrences of "*yeah*", out of which 92 were in utterance initial position and 18 occurred immediately after some non-content turn-preface items or immediately after a lexical repetition, also in initial position. The Brazilian subjects of this study employed "*yeah*" largely as a convergence token, signalling a tendency for agreement within preference structure. There were occasional occurrences as a continuer token and as an information receipt token. A number of occurrences showed a similar use to "*yes*" within answering moves to "*yes*" or "*no*" questions. Native speaker data showed a far higher frequency of "*yeah*" in utterance initial position, which corroborates the assumption that "*yeah*" is mostly used as a response token. The Brazilian learners of this study produced "*yeah*" less frequently than native speakers did but displayed a consistent use of "*yeah*" as a listener response token. Comparisons with the most frequent collocations of "*yeah*" as produced by native speakers showed that the subjetcs of this study employed "*yeah*" as a response token within similar contexts of usage.

The Brazilian corpus displayed 18 occurrences of "*uhuh*", out of which 17 were in utterance initial position and 1 was in medial position. 14 occurrences of "*uhuh*" appeared as one-word insert sequences in utterance initial position.There were only 2 occurrences of "*uhuh*" preceding extended responses and only 1 occasional occurrence preceding paralinguistic elements. While the Brazilian learners used "*uhuh*" more frequently as a convergence token, its context of usage varied. The second most common function was "*uhuh*" as a continuer token, entirely as one-word insert sequences. The least frequent function was as an information receipt token. Despite insufficient native speaker data, it was suggested that the Brazilian learners of this study employed "*uhuh*" as a one-word response token performing

the convergence, continuer and information receipt functions at relevant places in conversations.

The final part of this chapter discussed "*yeah*" and "*uhuh*" findings in the light of its specific research sub-question. It was indicated that the subjects of this study employed both "*yeah*" and "*uhuh*" more frequently as convergence tokens. However, they used "*yeah*" 12 times more frequently than "*uhuh*" for this function. Conversely, the Brazilian learners used the non-word vocalisation more frequently as a continuer token and as an information receipt token. It was suggested that the Brazilian learners of this study displayed a preference for the short utterance form to express convergence whereas they seemed to prefer the non-word vocalisation for speech management functions, namely continuer and information receipt. Furthermore, comparisons with data from young British women (O'Keeffe and Adolphs, 2008, p.90) indicate possible underuse of engagement, continuer and information response tokens by the Brazilian learners of this study.

As for pedagogical recommendations, it was concluded that the subjects of this study would benefit from classroom activities that offer learners the opportunity to employ more varied minimal and non- minimal response tokens in general. The analyses of the forms fitting the description of minimal and non -minimal forms showed a limited range. Learners would also gain from activities aiming at different forms to emotionally respond to previous communicative messages, to encourage speakers to continue their ongoing turns and to acknowledge understading. Following from this, activities based on Thornbury and Slade's (2006, p. 296) three part cycle "exposure, instruction and practice" were proposed.

The investigations of discourse marking adverbs, explicit and implicit adverbial hedges and minimal response tokens in the Brazilian learner corpus were intended to illustrate some of the key phenomena comprising conversational competence in the management of discourse, in the negotiation of illocutionary meaning and in conversational practices, respectively. Thus, it is possible to say that the final objective of this thesis has been accomplished. Let us round off this thesis by commenting on the impact these findings may have on proposed objectives.

CONCLUSION

The theme of this thesis was conversational competence in English as a second language**.** This study pursued three major aims: to revisit the construct "conversational competence" in L2 and place it within a pragmatic domain, grounded on the notion of "meaning in interaction" (Thomas, 1995); to detail key discourse, pragmatic and interactional phenomena which comprise conversational competence in L2; to investigate some of the key discourse, pragmatic and interactional phenomena comprising conversational competence in the oral production of twenty Brazilian learners of English at CEFR B1. As conversational competence is a broad concept encompassing a number of different features, pragmatic markers that seem to be prototypical of its proposed facets were chosen for investigations, namely discourse marking adverbs, adverbial hedges and backchannels.

Chapter 1 presented the foundations of Discourse Analysis, Pragmatics and Conversation Analysis and detailed discourse, pragmatic and interactional phenomena inherent in face-to-face conversations. A functional approach to discourse analysis (Brown & Yule, 1983) was chosen since it focuses on the analysis of language in use. It was highlighted that language performs both transactional and interactional functions, namely the communication of factual and propositional information and the establishment and maintenance of social relationships, respectively. Pragmatics as meaning in interaction (Thomas, 1995) was the perspective adopted in this study as making meaning is a dynamic process, involving the negotiation of meaning between the speaker and the hearer, the context of an utterance and its meaning potential. Conversation Analysis was introduced as a sociological approach that views ordinary conversation as a deeply ordered, structurally organised phenomenon and utterances as objects that speakers use to accomplish particular things in their interactions with others (Hutchby & Wooffitt, 2008).

Chapter 2 addressed Interlaguage Pragmatics and the second language learner. Leech (1983) was acknowledged to have first related pragmatics to second language acquisition. It was stated that Selinker's notion of interlanguage (1972) first formalised the interplay between SLA and pragmatics. Interlanguage Pragmatics was defined as a second-generation hybrid belonging to the interdisciplinary fields Second Language Acquisition and Pragmatics. It was argued that Canale and Swain (1980) and Canale's (1983) communicative competence model

failed to indicate how its different components interact with one another (Csépes, 2009). Likewise, while Bachman's original framework of communicative competence (1990) represents a major reconceptualisation since it locates competence within a wider performance framework (Csépes, 2009), it does not account for the co-constructed nature of communicative performance (McNamara, 1997). This thesis claimed that the ability to produce spoken discourse and interact with other speakers seems complex enough to justify a separate characterisation from the ability to produce written discourse, especially when one considers the negotiation of meaning that takes place in real time.

Following from this, a model for conversational competence within a pragmatic domain was proposed, grounded on two assumptions: conversation is prototypical of language usage (Levinson, 1983) and language in use involves the negotiation of meaning (Thomas, 1985). Conversational competence in a second language seems to comprise three major interrelated facets:

a) The management of discourse includes a learner's ability to:

- produce and sustain (longer) stretches of discourse, including the use of cohesive devices , co-reference, substitution, ellipsis and lexical relationships in order to convey cohesive relations;
- produce coherent discourse by adhering to thematic and information structure, which may be strengthened by cohesive relations;
- employ discourse markers to launch and conclude topics, to signal relationships of sequence, to monitor and manage the ongoing discourse, to indicate that they have not selected the most appropriate way of expressing things and that they are adding to or refining what they say, to signal that they are sensitive to listeners' needs and that they are monitoring the state of shared knowledge and to mark their stance or attitude towards the message.

b) The negotiation of illocutionary meaning includes a learner's ability to:

- produce illocutionary acts according to the intended illocutionary force, relying on both pragmalinguistic and sociopragmatic knowledge and conveying the desired degree of power and relative distance;
- employ positive and negative politeness strategies in order to minimise the degree of imposition of face-threatening acts and to produce face-saving acts;

- use pragmatic force modifiers in order to soften or strengthen the force of their messages and to make concepts fuzzier or less fuzzy;
- interpret the illocutionary force of utterances, including the understanding of indirect speech acts, based on pragmalinguistic and sociopragmatic knowledge, illocutionary force indicating devices and contextual clues.

c) The deployment of conversational practices includes a learner's ability to:

- carry out small talk as a means to initiate a conversation, including conversational routines;
- orient themselves to the rules of turn-taking;
- construct sequences of utterances by reference to the practices of adjacency pair organisation;
- show an orientation to preference structure;
- employ backchannels in order to indicate that they are following what is being said and also to inform how the message is being received.

The above model for conversational competence was then related to the CEFR and its level descriptors. The CEFR levels are based on six broad reference levels: A1 (elementary), A2 (pre-intermediate), B1 (intermediate), B2 (upper-intermediate), C1 (advanced) and C2 (proficiency). It was concluded that the different components of conversational competence are present as from level A1 and that the interplay of discourse, pragmatic and interactional phenomena is more noticeable as from level B1. The choice of CEFR level B1 for empirical investigation was then justified, based on the intermediate plateau phenomenon and on the requirements of the Brazilian programme "Science without Borders".

Chapter 3 focused on detailing the methodology and procedures employed in the empirical investigations. This study employed Corpus Linguistics as its methodogy and relied on form-to-function investigations. It included the participation of twenty Brazilian learners of English at CEFR B1+ who attended a general English course at a language institute in the South of Brazil, between the second semester of 2012 and the second semenster of 2013. A small corpus comprising 10 manually transcribed FCE interviews (public version) was built. The transcription included extra-linguistic information, filled pauses and backchannels. It was stressed that learners' oral production was reproduced as originally uttered and that no attempt

was made to correct their grammatical, lexical, phonological and pragmatic mistakes.The Corpus Query Processor web server (CQPweb) at Lancaster University was used for access to benchmark corpora and for native speaker quantitative and qualitative analyses. As the subjects of this investigation had been exposed to textbooks and classroom materials based mostly on the British variety, it felt methodologically adequate to focus on British English native speaker corpora available on CQPweb: the BNC Sampler spoken sub-corpus and a specialised corpus of "face-to-face conversations from the 1990s" compiled from the Diacronic Corpus of Present-day Spoken English. In addition to these corpora, other native speaker and learner corpora were referred to in chapter sub-sections for specific purposes. As CQPweb does not allow users to upload their own corpora, the free software "AntConc 3.4.1" (Anthony, 2014) was used for learner data analyses. Word List, Concordance and File View were the tools used in the analyses of the Brazilian learner corpus.

Chapter 4 investigated the most common discourse marking adverbs used to mediate segments of discourse in conversations. The Brazilian learners of this study produced the following most frequent adverbs which seem to fit the characterisaton of discourse markers: "*maybe*", "*really*", "*just*", "*well*", "*probably*", "*actually*". The adverbs "*really*", *"well*" and "*actually*" were chosen for investigation in chapter 4 since they are commonly used to mediate segments of discourse with regards to the different parts of discourse and also to the relations between speakers and parts of discourse. The comparison of relative frequencies with native speaker data indicated considerable underuse of the discourse markers "*well*" and "*actually*" and a consistent use of "*really*" by the subjects of this study.

All occurrences of "*well*" as a one-word adverb portrayed its use as a discourse marker, mediating segments of discourse: choice-related (for pausing and planning what to say next), change-related (for repair) and opinion (for conveying stance). Subjects employed "*well*" mostly for speech management functions. Attitudinal functions were less frequent. These data are aligned with Aijmer's learner results (2011), corroborating the assumption that learners use "*well*" more frequently for speech management functions than native speakers do (LOCNEC data) and less for attitudinal ones. The subjects of this study used "*well*" more frequently in the choice-related function than in the change-related one, in a similar way as the Swedish learners of LINDSEI, differing from native speaker patterns. "*Well*" was also

found surrounded by filled and unfilled pauses, strengthening Aijmer's claim (2011) that learners may use it more as a hesitation device rather than as a marker of deliberation.

The Brazilian learners of this study were able to produce "*really*" so as to perform all the pragmatic functions proposed by Paradis (2003), including the de-emphasiser one, though overall relative frequencies differed. Subjects used "*really*" mostly as a degree reinforcer of scalar adjectives and as an emphasiser accompanying attitudinal verbs. All scalar adjectives collocating with "*really*" showed a positive connotation in their actual realisation. As for discourse properties, truth attester "*really*" was observed in utterance initial and final positions mediating segments of discourse, but with a lower frequency.This difference signals that the subjects of this study produced "*really*" more frequently within utterances, either to emphasise scalar properties of adjectives or to reinforce the subjectivity of situations when accompanying attitudinal verbs. It possibly also indicates less familiarity with "*really*" truth attester, which takes scope over full propositions.

The Brazilian corpus displayed "*actually*" mostly as a discourse particle since most occurrences were placed in utterance initial position. All occurrences were followed by "*I*" and verbs expressing preferences. In the examples of the discourse particle, "*actually*" was used to introduce one's opinion or preference conveying different attitudes to the preceding discourse.The single occurrence of "*actually*" in medial position showed its use as a mitigating device. The low frequencies of "*actually*" in the Brazilian data indicate learners' underuse of the discourse marker and of the regular adverb. Furthermore, the absence of "*actually*" in utterance final position may signal lack of familiarity with its functions as a floor holder, a marker of subjective opinion or an afterthought.

As for the distribution of forms among individual learners, 65% of the subjects of this study produced at least one form of discourse marking adverbs used to mediate segments of discourse. While "*really*" was used by 60% of the subjects, "*well*" and "*actually*" were employed by only 15% of the subjects. In addition, 25% of the subjects produced two forms of discourse marking adverbs, showing a more consistent command. One subject of this study overused "*really*". 35% of the subjects did not produce any discourse marking adverb. These results attest learners' limited command of discourse markers in general.

Chapter 5 focused on the most common explicit and implicit adverbial hedges used to mitigate representative speech acts. The Brazilian learner corpus displayed 18 lexical items (or lemmas) which fit the form characterisation of hedges as proposed by O'Keeffe, McCarthy and Carter (2007, p.175) and/or are traditionally associated with mitigation. Nevertheless, not all forms were necessarily employed for mitigating purposes. The adverbial hedging forms "*maybe*", "*probably*" and "*just*" were found mostly within contexts of representative speech acts such as asserting, making deductions and concluding. They were chosen for the empirical investigations of explicit and implicit hedges.The Brazilian learners of this study used "*maybe*" more frequently than "*probably*", differing from British native speaker linguistic behaviour and aligned with American English patterns (Biber et al., 1999). Subjects displayed a considerably higher relative frequency of "*maybe*" than those in native speaker corpora, signalling overuse. Comparisons of relative frequencies attested an extremely high frequency of "*just*" in both benchmark corpora and indicated low figures of "*just*" in the learner corpus, signalling underuse.

Both epistemic adverbs "*probably*" and "*maybe*" were produced mostly in medial position, despite differences in individual relative frequencies. The second most frequent position was "*maybe*" utterance initial. Both adverbial forms were employed as explicit hedges to mitigate representative speech acts. All occurrences of "*probably*" displayed its use as an explicit adverbial hedge used to mitigate assertions, either mitigating the degree of certainty of propositions (when speculating about visual prompts) or marking uncertainty about the effectiveness of proposed ideas (reaching an agreement) . 94.38 % of the occurrences of "*maybe*" also showed its use as an explicit adverbial hedge employed to mitigate assertions, mitigating the degree of certainty of propositions, marking uncertainty when discussing pros and cons of suggested ideas, softening individual preferences and as a politeness hedge when expressing disagreement with proposed ideas. 2.81% of the occurrences of "*maybe*" demonstrated its use mitigating another speech act category, namely directives (suggestions). 2.81% of its occurrences displayed its use as a hesitation filler.

The Brazilian learners of this study produced "*just*" mostly in medial position. "*Just*" was more frequently used in medial position to emphasise elements within utterances as a means to modify the pragmatic force of utterances (mitigating). The second most common function was the use of "*just*" as a restrictive adverb, taking scope over the truth condition of particular

elements within the utterance. Therefore, the empathic function of "*just*" seems to be the unmarked use by the Brazilian learners of this study, similarly to Aijmer's findings (2002) in native speaker corpora. Nonetheless, the percentage difference between the emphatic and restrictive functions in the Brazilian corpus is much lower, possibly signalling a stronger tendency to use "*just*" as a restrictive adverb as well. The effect of "*just*" on the pragmatic force of utterances produced by the Brazilian learners was sometimes fuzzy, even within its contexts.

80% of the Brazilian learners of this study employed at least one explicit adverbial hedge to mark epistemic stance. While 75% of the subjects produced "*maybe*", "*probably*" occurred in 30% of individual learners' production. 40% of the subjects clearly overused "*maybe*" (4 or more occurrences). In addition, 25% of the Brazilian learners employed two forms of explicit adverbial hedges to convey epistemic stance. 45% of the subjects of this study produced "*just*" either as an implicit adverbial hedge or as a restrictive adverb. Overall, only 25% of the subjects employed one form of explicit adverbial hedges and "*just*" (irrespective of functions). 10% of the Brazilian learners were able to produce the three adverbial hedging forms under investigation. One can conclude that most subjects of this study were able to employ explicit adverbial forms to hedge representative speech acts while "*just*" as an implicit modifier or as a restrictive adverb posed more problems to learners.

Chapter 6 examined the most common minimal response tokens used to express good listenership. The Brazilian learner corpus displayed the following most frequent tokens fitting the description of minimal forms (O'Keeffe and Adolphs, 2008): "*uh*", "*yeah*", "*hum*", "*ah*", "*uhuh*", "*oh*" and "*aih*". Nonetheless, not all occurrences of individual forms were employed as response tokens in utterance initial position or as one-word insert sequences. A closer look at the contexts of usage of individual forms revealed that "*yeah*" was the only short utterance produced by the subjects of this study as a response form. "*Uhuh*" was the non-word vocalisation most frequently used as a response token. Therefore, the minimal forms "*yeah*" and "*uhuh*" were chosen for the empirical investigations.The Brazilian learners of this study produced "*yeah*" less frequently than native speakers did but they displayed a consistent use of "*yeah*" as a listener response token in utterance initial position. Due to insufficient native speaker corpus evidence of "*uhuh*", its relative frequency was compared to that of more

advanced learners (Spanish component of LINDSEI). The Brazilian learners of this study produced "*uhuh*" slightly less frequently than the Spanish learners did.

The Brazilian subjects of this study employed "*yeah*" largely as a convergence token (88.19%), signalling a tendency for agreement within preference structure. Nonetheless, some occurrences of "*yeah*" as a convergence token displayed a contradiction, as in the cases of "*yeah*" followed by attitudinal verbs in negative forms or by "*but*". In those examples, the token "*yeah*" seems to have been employed as an informal politeness formula to reinforce common ground despite disagreement. There were occasional occurrences of "*yeah*" as a continuer token (1.82%) and as an information receipt token (0.90%). As for other observed functions, a number of occurrences (8.19%) showed a similar use to "*yes*" within answering moves to "*yes*" or "*no*" questions and as a compensation strategy when searching for words (0.90%).

The subjects of this study employed "*uhuh*" in utterance initial position largely as one-word inserts (82.35%). Furthermore, the qualitative analyses of "*uhuh*" indicated 3 response functions, namely as a convergence token, as a continuer token and as an information receipt token. While the Brazilian learners used "*uhuh*" more frequently as a convergence token, its context of usage varied. The second most common function was "*uhuh*" as a continuer token, entirely as one-word insert sequences. The least frequent use was as an information receipt token, both as a one-word insert sequence or preceding an extended response.

Comparing learners' preference for the minimal response forms, subjects used the short utterance "*yeah*" twelve times more frequently than the non-word vocalisation "*uhuh*" in the convergence function. Learners seemed to prefer the non-word vocalisation "*uhuh*" for speech management functions, namely continuer and information receipt functions. 80% of the subjects of this study produced the short utterance "*yeah*". 40% of the Brazilian learners employed both "*yeah*" and "*uhuh*" to express good listenership. "*Uhuh*" was never produced in isolation as a response token by these subjects, who also employed "*yeah*" with much higher frequencies, with the exception of one subject. These results indicate a consistent use of "*yeah*" as a response token by the Brazilian learners of this study.

Comparing the distribution of the minimal forms among response functions, convergence tokens occurred 13 times as frequently as continuer tokens and 26 times as frequently as information receipt tokens. The minimal forms were not employed to express engagement. Thus, a further investigation was then manually carried out in order to identify non-minimal forms of engagement tokens in the Brazilian corpus. Results indicated possible underuse of engagement, continuer and information response tokens. However, these findings are to be considered partial since a comprehensive account of response tokens and functions would require the analyses of all the minimal and non-minimal forms in the Brazilian corpus.

All in all, the results of the empirical investigations highlighted some areas which seem to have posed more difficulties to Brazilian learners at CEFR B1. Undoubtedly, the subjects of this study would benefit from some pedagogical assistance for acquiring discourse markers in general, based on the limited range of discourse markers observed in their spoken production and on the underuse of the pragmatic and discourse functions of "*well*" and "*actually*". The Brazilian learners of this study produced the discourse marker "*well*" mostly as a hesitating device and with a considerably lower frequency than that of native speakers. Few occurrences of attitudinal "*well*" were observed. As for "*actually*", only 4 occurrences were observed in the Brazilian corpus and none of which had been placed in utterance final, its most common position as a discourse marker in native speaker production.

Another pedagogical focus would be exposure to other forms to convey epistemic stance, based on the overuse of "*maybe*" and on the absence of "*may*" and "*might*" in the Brazilian data. Learners would also benefit from more exposure to the pragmatic functions of "*just*", due to relatively low figures observed in the Brazilian data.Within EFL contexts, the modal verbs "*must*", "*may*", "*might*", "*could*" and "*can't*" are usually grouped as "modal verbs for deductions" and introduced to learners in general as from CEFR A2. The pragmatic functions of "*just*", on the other hand, are usually presented to students at more advanced levels, usually at CEFR B2 or C1.

A third pedagogical objective would be exposure to classroom activities that offer learners the opportunity to employ varied minimal and non- minimal response tokens in general. The analyses of the forms fitting the description of response tokens showed a limited range in the Brazilian data. In addition, the subjects of this study did not produce engagement, continuer

and information receipt tokens in the same proportion among functions as native speakers did. Thus, learners would also gain from activities aiming at different forms to emotionally respond to previous communicative messages, to encourage speakers to continue their ongoing turns and to acknowledge understanding.

As proposed in section 2.3.2, conversational competence in L2 within a pragmatic domain encompasses the management of discourse (mediating segments of discourse), the negotiation of illocutionary meaning (producing, manipulating and interpreting the pragmatic force of illocutionary acts) and the deployment of conversational practices (showing an orientation to turn and sequence organisation and preference structure). Thus, the empirical investigations of this thesis revealed only a small portion of learners' oral production in each domain. Some elements under investigation displayed low raw frequencies in the Brazilian learner corpus. As a consequence, my comments can only be indicative of general patterns. Nonetheless, the qualitative analyses allowed us to see the deployment of discourse marking adverbs, explicit and implicit hedges and minimal response tokens to convey different shades of pragmatic meaning by the Brazilian learners of this study.

As for the intermediate plateau phenomenon, the qualitative analyses of this study showed that some subjects displayed a limited vocabulary range and persistent fossilised language errors. On some occasions, nevertheless, learners' oral production sounded more natural or fluent mainly because of their use of pragmatic markers. Thus, one can suggest that a programme aimed at helping learners move beyond the intermediate plateau should provide exposure to, instruction on and practice of both code and use features. Suggested elements include grammatical, lexical and phonological features, functional exponents, fixed expressions and conversational routines employed to perform language functions and specific lexical items, lexical bundles and pragmatic markers used to convey different shades of pragmatic meaning. The pedagogical recommendations proposed in the final sections of chapters 4, 5, and 6 may be of some value for teachers whose students display similar difficulties. Furthermore, the descriptions of the three facets of conversational competence list the functions of key discourse, pragmatic and interactional phenomena in conversations. Teachers may also consider these descriptions when deciding on a syllabus for a programme focusing on speaking sub-skills.

From a theoretical perspective, the investigations of specific items illustrated the wide range of functions that pragmatic markers can perform, which in turn, indicates that the same form can operate in the different facets of conversational competence for different purposes. In the Brazilian learner data, the pragmatic marker "*actually"* was used as a discourse marking adverb to introduce one's opinion or preference conveying different attitudes to the preceding discourse. However, it was also employed as a mitigating device. The epistemic adverb "*maybe*" was employed as an explicit adverbial hedge to mitigate representative and directive speech acts. By contrast, learners also used it for planning ahead, as a hesitation filler. The short utterance "*yeah*" was produced as a listener response token and also as a compensation strategy when searching for words. Therefore, one can assume that the qualitative analyses of this study provide some evidence of the overlapping functions of pragmatic markers. They also corroborate the interplay between discourse, pragmatic and interactional phenomena in the different facets of conversational competence.

As for future directions, the overuse of "*maybe*" in the Brazilian data offers a rich opportunity for further investigations within Interlanguage Pragmatics. Is the observed tendency to use explicit adverbial hedges to convey epistemic stance instead of modal verb forms due to the fact that learners have not yet acquired the epistemic use of "*may*" and "*might*"? Could this tendency be also interpreted as negative transfer? Longitudinal learner investigations combined with cross-cultural corpus-based studies may provide some insight into this apparent overuse of "*maybe*" in the Brazilian learner corpus.

Furthermore, the Brazilian learner data were retrieved from a small spoken corpus comprising 10 FCE interviews. As a result, a few items displayed low raw frequencies. Thus, further Brazilian learner spoken data at CEFR B1 will be collected in the future. A larger and more representative corpus will hopefully corroborate the assumption that a limited command of discourse markers and a limited range of response forms are prototypical characteristics of Brazilian learners' oral production at CEFR B1 level.

Lastly, this study investigated some of the key discourse, pragmatic and interactional phenomena comprising conversational competence. Future investigations focusing on other features of conversational competence are required so as to present a broader view of conversational competence at CEFR B1.

BIBLIOGRAPHICAL REFERENCES

Adolphs, S. (2008). *Corpus and Context: Investigating Pragmatic Functions in Spoken Discourse.* Amsterdam: John Benjamins Publishing Company.

Aijmer, K. (1996). *Conversational Routines in English: Convention and Creativity.* London: Longman.

Aijmer. K. (2002). *English Discourse Particles: Evidence from a Corpus.* Amsterdam: John Benjamins.

Aijmer. K. (2008). At the interface between grammar and discourse: A corpus-based study of some pragmatic markers. In: Romero-Trillo, J. (Ed). *Pragmatics and Corpus Linguistics*: *A Mutualistic Entente* (pp. 11-36). Berlin: Mouton de Gruyter.

Aijmer. K. (2009). Interjections in the COLT Corpus. In: Slembrouck, S.; Taverniers, M.; Herreweghe, M. (Eds.) *From Will to Well: Studies offered to Anne-Marie Simon-Vanderbergen* (pp. 11-19). Gent: Academia Press.

Aijmer. K. (2011). Well I'm Not Sure I Think…: The Use of Well by Non-native Speakers. *International Journal of Corpus Linguistics, Volume 16 (2),* pp. 231-254.

Allwood, J., Andersson, L., Dahl, O. (1977*). Logic in Linguistics.* Cambridge: Cambridge University Press.

Andersen, G. (2001). *Pragmatic Markers and Sociolinguistic Variation: A Relevance-Theoretic Approach to the Language of Adolescents.* Amsterdam: John Benjamins.

Anthony, L. (2014). AntConc (Version 3.4.2) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/.

Austin, J. (1962). *How to Do Things with Words***.** Oxford: Oxford University Press.

Austin, J. (1975). *How to Do Things with Words* (2nd Edition)**.** Harvard: Harvard University Press.

Bach, K., Harnish, R. M. (1979). *Linguistic Communication and Speech Acts.* Cambridge, MA: MIT Press.

Bachman, L. (1990). *Fundamental Considerations in Language Testing.* Oxford: Oxford University Press.

Bachman, L., Palmer, A. (1982). The Construct Validation of Some Components of communicative proficiency. *TESOL Quarterly, number 16, Volume 4,* pp. 449-465.

Bachman, L., Palmer, A. (1996). *Language Testing in Practice.* Oxford: Oxford University Press.

Baker, P. (2006). *Using Corpora in Discourse Analysis.* London: Continuum International Publishing Group.

Ballmer, T., Brennenstuhl, W. (1981). *Speech Act Classification: A Study in Lexical Analysis of English Speech Activity Verbs.* Berlin: Springer-Verlag.

Bialystok, E. (1993). Symbolic Representation and Attentional Control in Pragmatic Competence. In: Kasper, G.; Blum-Kulka (Eds.) *Interlanguage Pragmatics* (pp. 43-57). Oxford: Oxford University Press.

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

Biber, D., Johansson, S., Leech, L., Conrad, S., Finegan, E. (1999*). Longman Grammar of Spoken and Written English.* Harlow: Pearson Education.

Blum-Kulka, S. (1991). Interlanguage Pragmatics: The case of requests. In: Phillipson, R.; Kellerman, E.; Selinker, M; Sharwood Smith, M.; Swain, M. (Eds.) *Foreign/second language pedagody research.* Clevedon and Philadelphia: Multilingual Matters.

Blum-Kulka, S., House, J., Kasper, G. (Eds.) (1989). *Cross-Cultural Pragmatics: Requests and Apologies [Advances in Discourse Processes], Vol. XXXI.* Norwood: Ablex.

Blum-Kulka, S., Olshtain, E. (1986). Too many words. Length of utterance and pragmatic failure. *Studies in Second Language Acquisition, Volume 8*, pp. 47-61.

Blum-Kulka, S., Sheffer, H. (1993). The Metapragmatic Discourse of American-Israeli Families at Dinner. In: Kasper, G.; Blum-Kulka (Eds.) *Interlanguage Pragmatics* (pp. 196-223). Oxford: Oxford University Press.

Bolinger, D. (1972) *Degree words.* Mouton.

Brinton, L.J. (1996) *Pragmatic Markers in English: Grammaticalization and Discourse Functions*. Berlin: Mouton de Gruyter.

British International Corpus of English (ICE). Distributed by University College London. URL: http://www.ucl.ac.uk/english-usage/projects/ice-gb/.

Brown, G., Yule, G. (1983). *Discourse Analysis.* Cambridge: Cambridge University Press.

Brown, P., Levinson, S. (1978). Universals in Language Usage: Politeness Phenomena. In: Goody, E.N. (Ed.) *Questions and politeness: Strategies in social interaction* (pp. 56-289). Cambridge: Cambridge University Press.

Brown, P., Levinson, S. (1987). *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.

Caffi, C. (2013). Mitigation. In: Sbisà, M.; Turner, K. *Handbook of pragmatics, Volume 2, Pragmatics of Speech Actions* (pp. 257-285). De Gruyter.

Canale, M. (1983). On some Dimensions on Language Proficiency. In: Oller, J.W. (Ed.). *Issues in Language Testing Research* (pp. 333-342). Rowley: Newbury House.

Canale, M., Swain, M. (1980). Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics 1*, pp. 1-47.

Carston, R. (2004). Explicature and Semantics. In: Davis, S. & Gillon, B. (Eds.) *Semantics: A Reader* (pp. 1-34). Oxford: Oxford University Press.

Carter, R., McCarthy, M. (1995). Grammar and the Spoken Language. *Applied Linguistics. Volume 16 (2)*, pp. 141-158.

Carter, R., McCarthy, M. (2006). *The Cambridge Grammar of English: A Comprehensive Guide to Grammar and Usage.* Cambridge: Cambridge University Press.

Celce-Murcia, M., Olshtain, E. (2000). *Discourse and Context in Language Teaching: a Guide for Language Teachers.* Cambridge: Cambridge University Press.

Chafe, W. (1986). Evidentiality in English Conversation and Academic Writing. In: Chafe, W.; Nichols, J. (Eds) *Evidentiality: The Linguist Coding of Epistemology* (pp. 261-272). Alex Publishing Corporation.

Conrad, S., Biber, D. (2000). Adverbial Marking of Stance in Speech and Writing. In: Huston, S.; Thompson, G. (Eds.) *Evaluation in Text: Authorial Stance and the Construction of Discourse* (pp. 56-73). Oxford: Oxford University Press.

Corpus Query Processor at Lancaster. Retrieved from https://cqpweb.lancs.ac.uk/.

Corsetti, C.R. (2009). *The Enhancement of Pragmatic Competencies via Listening Activities.* (Unpublished Master's dissertation). Pontifícia Universidade do Rio Grande do Sul, Porto Alegre.

Council of Europe (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.

Coulmas, F. (1981). Position to your soul: Thanks and apologies contrastively viewed. In: Coulmas, F. (Ed.) *Conversational routine: Explorations in standardized communication situations and pre-patterned speech*, pp. 69-91. The Hague: Mouton.

Cowley, A. (2012). *World Wise 1*. Rio de Janeiro: Learning Factory.

Csépes, I. (2009). Measuring Oral Proficiency through Paired-Task Performance. In: Grotjahn, R ; Sigott, G. *Language Testing and Evaluation, Volume 14*. Peter Lang.

Culpeper, J. (1996). Towards an anatomy of impoliteness. *Journal of Pragmatics 25*, pp. 349-367.

Culpeper, J. (2005). Impoliteness and entertainment in the television quiz show: The Weakest Link. *Journal of Politeness Research: Language, Behaviour, Culture 1*, pp. 35-72.

Culpeper, J. (2011). *Impoliteness: Using Language to Cause Offence*. Cambridge: Cambridge University Press.

Culpeper, J., Bousfield, D., Wichmann, A. (2003). Impoliteness revisited: With special reference to dynamic and prosodic aspects. *Journal of Pragmatics* 35, pp. 1545-1579.

Culpeper, J., Haugh, M. (2014). *Pragmatics and the English Language.* Palgrave Macmillan.

Culpeper, J., Kytö, M. (2010). *Early Modern English Dialogues: Spoken Interaction as Writing.* Cambridge: Cambridge University Press.

Evinson, J. (2010). What are the basics of analysing a corpus? In: O'Keefe, A; McCarthy, M. *The Routledge Handbook of Corpus Linguistics*, pp. 122-135. Routledge.

Fairclough, N. (1995). *Critical Discourse Analysis*. Boston: Addison Wesley.

Fraser, B. (1996). Pragmatic Markers. *International Pragmatics Association*, 6:2, pp. 167-190.

Galaczi, E. (2008). Peer-Peer Interaction in a Speaking Test: The Case of the First Certificate in English Examination. *Language Assessment Quartely*, 5:2, pp. 89-119.

Galaczi, E., ffrench, A. (2011). Context Validity. In: Taylor, L. *Examining Speaking: Research and Practice in Assessing Second Language Speaking, Studies in Language Testing 30,* pp. 112-170. Cambridge: UCLES/Cambridge University Press.

Gass, S. M., Selinker, L. (2008). *Second Language Acquisition: An Introductory Course* (3rd Edition). Routledge.

Gilbert, S., Staub, G. (2014) Examiner confidence survey: An investigation into Speaking Examiners´ confidence in the accuracy of the assessments they make. Cambridge English Research Notes 57, pp.50-59. UCLES.

Granger, S. (2002). A Bird's eye view of learner corpus research. In: Granger, S., Hung, J., Petch-Tyson, S. (Eds.). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, pp. 3-33. John Benjamins Publishing Company.

Graig, R., Tracy, K., Spisak, F. (1986). The discourse of requests: Assessment of a politeness approach. *Human Communication Research, Volume 12*, pp. 437-486.

Grice, E, P. (1975). Logic and Conversation. In Cole, P.; Morgan, J. (Eds.). *Syntax and Semantics, Volume 3,* pp. 41-58. New York: Academic Press.

Grice, E, P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Gumperz, J.J. (1982). *Discourse strategies*. Cambridge: Cambridge University Press.

Halliday, M. A. K. (1985). *An Introduction to Functional Grammar*. Edward Arnold.

Halliday, M. A. K., Hassan, R. (1976). *Cohesion in English*. London: Longman.

Hardie, A. (2012). CQPweb: combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics 17:3*, pp. 380-409.

Harmer, J. (2012). *Essential Teacher Knowledge: Core Concepts in English Language Teaching*. Pearson Education Limited.

Hawkings, J. A., Filipovic, L. (2012). *Critirial Features in L2 English: Specifying the Reference Levels of the Common Euroepean Framework*. Cambridge: Cambridge University Press.

Heritage, J., Atkinson, J. M. (1984). Introduction. In: Heritage, J.; Atkinson, J.M. (Eds.). *Structures of Social Action*, pp. 1-15. New York: Cambridge University Press.

Hölker, K. (1991). Französisch: Partikelforschung. *Lexikon der Romaninstischen Linguistik, V*, pp. 77-88. Tübingen: Niemeyer.

Huston, S. (2002). *Corpora in Applied Linguistics.* Cambridge: Cambridge University Press.

Hutchby, I., Woodffitt, R. (2008). *Conversation Analysis* (2nd Edition). Cambridge: Polity Press.

Hymes, D. (1972). On Communicative Competence. In: Pride, J.P.; Holmes, J. (Eds.) *Sociolinguistics: Selected Readings*, pp. 269-293. Harmondsworth: Penguin.

Jefferson, G. (1986). Notes on the latency in overlap onset. *Human Studies, 9*, pp. 153-183.

Jeffery, P., Lloyd, M., Goldstein, B. (2009). *New Framework 4*. Richmond.

Jucker, A. (1993). The discourse marker well: A relevance-theoretical account. *Journal of Pragmatics, V19 (5)*, pp. 435-452.

Jucker, A. (2013). Corpus Pragmatics. In: Ostman, J; Verschueren, J. *Handbook of Pragmatics: 2013 Installment*, pp. 1-17. John Benjamins Publishing Company.

Jucker, A., Taavitsain, I. (2014). Complimenting in the history of American English: A metacommunicative expression analysis. In: Taavitsain, I., Jucker, A., Tuominen. (Eds.) *Diachronic Corpus Pragmatics*, pp. 257-276. Amsterdam: John Benjamins.

Jucker, A., Ziv, Y. (1998). Discourse Markers: Introduction. In: Jucker, A., Ziv, Y. (Eds.) *Discourse Markers: Descriptions and Theory.* John Benjamins.

Kasper, G. (1997). *Can Pragmatic Competence Be Taught?* University of Hawaii.

Kasper, G., Blum-Kulka, S. (1993). Interlanguage Pragmatics: An Introduction. In: Kasper, G., Blum-Kulka (Eds.). *Interlanguage Pragmatics*, pp. 3-17. Oxford University Press.

Kasper, G., Rose, K. (2002). *Pragmatic Development in a Second Language.* Oxford: Blackwell Publishing.

Kilgarriff, A., Rychly, P., Smrz, P., Tugwell, D. (2014). The Sketch Engine: ten year on. In *Lexicography*, pp. 1-30.

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

Lakoff, G. (1972). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Papers from the Eight Regional Meeting of Chicago Linguistic Society*, pp. 183-228.

Lee, D. Y. W. (2010). What corpora are available? In: O'Keefe, A; McCarthy, M. *The Routledge Handbook of Corpus Linguistics*, pp. 107-121. Routledge.

Leech, G. (1983). *Principles of Pragmatics.* London: Longman.

Leech, G. (2000). Grammars of spoken English: new outcomes of corpus-orientated research. *Language Learning* 50, *Volume 4*, pp. 675-724.

Levinson, S. C. (1983). *Pragmatics.* Cambridge University Press.

Levinson, S. C. (2000). *Presumptive Meanings: the Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.

Manes, J., Wolfson, N. (1981). The Compliment formula. In: Coulmas, F. (Ed.) *Conversational Routine: Explorations in Standardized Communication Situations and Prepatterned Speech*, pp. 115-132. The Hague: Mouton.

McCarthy, M. (2002). Good listenership made plain: British and American non-minimal response tokens in everyday conversation. In: Reppen, R., Fitzmaurice, S., Biber, D. *Using Corpora to Explore Linguistic Variation*. Amsterdam: John Benjamins.

McCarthy, M. (2003). Talking Back: "Small" Interactional Response Tokens in Everyday Conversation. *Research on Language and Social Interaction 36 (1)*, pp. 33-63.

McEnery, T., Hardie, A. (2012). *Corpus Linguistics: method, theory and practice*. Cambridge: Cambridge University Press.

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

McKay, S. (2002). *Teaching English as an International Language.* Oxford: Oxford University Press.

McNamara, T. F. (1997). 'Interaction' in second language performance assessment: Whose performance? *Applied Linguistics, 18*, pp. 446-466.

McNamara, T. F. (2000). *Language Testing.* Oxford: Oxford University Press.

Mendelsohn, D. (1995). Applying Listening Strategies in the Second/Foreign Language Listening Comprehension Lesson. In: Mendelsohn, D., Rubin, J. (Eds.). *A Guide for the Teaching of Second Language Listening.* San Diego, CA: Dominie Press.

Mendelsohn, D. (1998). Teaching Listening. *Annual Review of Applied Linguistics.*

Milanovic, M., Weir, C. J. (2010). Series Editors' note. In: Martyniuk, W. *Aligning Tests with the CEFR*: *Reflections on using the Council of Europe 's draft manual. Studies in Language Testing 33*, pp. viii-xx. Cambridge: UCLES/Cambridge University Press.

Morris, C. (1938). Foundations of the Theory of Signs. In: Neurath, O., Carnap, C., Morris, C. (Eds.). *International Encyclopedia of Unified Science, Volume 1, Number 2,* pp. 13-71. Chicago: University of Chicago Press.

Mott, H., Petrie, H. (1995). Work place interactions: Women's linguistic behaviour. *Journal of Social Psychology 14*, pp. 324-336.

Nikula, T. (1992). *Lexical certainty modifiers in non-native and native discourse* (Unpublished licentiate thesis). University of Jyväskylä, Jyväskylä.

Nikula, T. (1996). *Pragmatic Force Modifiers: A Study in Interlanguage Pragmatics.* Jyväskylä: University of Jyväskylä.

North, B., Martyniuk, W., Panthier, J. (2010). Introduction: The Manual for Relating examinations to the Common European Framework of Reference for Languages in the context of the Council of Europe's work on language education. In: Martyniuk, W. *Aligning Tests with the CEFR: Reflections on using the Council of Europe 's draft manual. Studies in Language Testing 33*, pp. 1-17. Cambridge: UCLES/Cambridge University Press.

Niezgoda, K., Röver, C. (2001). Pragmatic and Grammatical Awareness. In: Rose, K., Kasper. G. (Eds.). *Pragmatics in Language Teaching,* pp. 63-79. Cambridge: Cambridge University Press.

O'Keeffe, A., Adolphs, S. (2008). Response tokens in British and Irish discourse: Corpus, context and variational pragmatics. In: Schneider, K.; Barron, A. *Variational Pragmatics: a focus on regional varieties in pluricentric languages*, pp. 69-98. John Benjamins Publishing Company.

O'Keeffe, A., McCarthy, M., Carter, R. (2007). *From Corpus to Classroom: language use and language teaching*. Cambridge: Cambridge University Press.

O'Keeffe, A., Walsh, S. (2012). Applying corpus linguistics and conversation analysis in the investigation of small group teaching in higher education. In: *Corpus Linguistics and Linguistic Theory, Vol. 8 (1),* pp. 159-181.

Palmer, F. R. (2001). *Mood and Modality* (2nd Edition). Cambridge: Cambridge University Press.

Paquot, M., Granger, S. (2012). Formulaic Language in Learner Corpora. *Annual Review of Applied Linguistics, 32*, pp. 130-149. Cambridge: Cambridge University Press.

Paradis, C. (2003). Between epistemic modality and degree: The case of *really.* In: Facchinetti, R., Krug, M., Palmer, F. *Modality in Contemporary English*, pp. 191-220. Mouton de Gruyter.

Perna, C. L. (2002). A Competência Pragmática na Realização de Pedidos de Desculpas em Inglês como L2. In: *Na Interface Semântica/Pragmática,* pp. 175-208. Porto Alegre: Edipucrs.

Pomerantz, A. (1984). Agreeing and Disagreeing with Assessments: Some Features of Preferred/Dispreferred Turn Shapes. In: Atkinson, J.M.; Heritage, J. (Eds.). *Structures of Social Action*: *Studies in Conversation Analysis*, pp. 57-101. Cambridge: Cambridge University Press.

Prince, E. F., Frader, J., Bosk, C. (1982). On hedging in the physician discourse. In: Pietro, R. J. D. (Ed) *Linguistics and Professions*, pp. 83-97. Norwood: Ablex.

Richards, J. C. (2006). Developing Classroom Speaking Activities: from Theory to Practice. *RELC Guideline**s**, Volume 28, number 2*, pp. 1-10. RELC, Singapore.

Richards, J. C. (2008). *Moving Beyond the Plateau: From Intermediate to Advanced Levels in Language Learning*. Cambridge: Cambridge University Press.

Richards, J. C., Sukwiwat, M. (1985). Cross-cultural Aspects of Conversational Competence. In: Richards, J. C. *The Context of Language Teaching,* pp. 128-143. Cambridge: Cambridge University Press.

Richards, J., Platt, J., Weber, H. (1990). *Longman Dictionary of Applied Linguistics*. Longman.

Röver, C. (2005). *Testing ESL Pragmatics.* Peter Lang GmBh.

Sacks, H. (1992). *Lectures on Conversation*. *Volume I.* Edited by G. Jefferson with Introduction by E.A. Schegloff. Oxford: Blackwell.

Sacks, H., Schegloff, E. A., Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking in Conversation. *Language. 50, 4*, pp. 696-735.

Sardinha, T.B. (2004). *Linguística de Corpus.* Barueri: Manole.

Scarcella, R. (1979). On speaking politely in a second language. In: Yorio, C. A., Perkins, K., Schachter, J. (Eds.). *On TESOL'79*, pp. 275-287. Washington: TESOL.

Scarcella, R., Oxford, R. (1992). *The Tapestry of Language Learning: the Individual in the Communicative Classroom*. Boston: Heinle & Heinle Publishers.

Schauer, G. A. (2009). *Interlanguage Pragmatic Development: The Study Abroad Context*. Continuum International Publishing Group.

Schauer, G., Adolphs, S. (2006). Expressions of gratitude in corpus and DTC data: vocabulary, formulaic sequences, and pedagogy. *System: An International Journal of Educational Technology and Applied Linguistics, Volume 34 (1),* pp. 119-134.

Schegloff, E. (1982). Discourse as an interacional achievement: some uses of "uh huh" and other things that come between sentences. In: Tannen, D (Ed.) *Analyzing Discourse: Text and Talk*, pp. 71-93. Washington: Georgetown University Press.

Schegloff, E. (2007). *Sequence Organization in Interaction: A Primer in Conversation Analysis, Volume 1*. Cambridge: Cambridge University Press.

Schegloff, E. A., Jefferson, G., Sacks, H. (1977). The Preference for self-correction in the organisation of repair in conversation. *Language, 53*, pp. 361-382.

Schegloff, E. A., Sacks, H. (1973) Opening up Closings. *Semiotica,7.4*, pp. 289-327.

Schmidt, R. (1983). Interaction, acculturation and acquisition of communicative competence. In: Wolfson, N., Judd, E. (Eds.). *Sociolinguistics and second language acquisition*, pp. 137-174. Rowley: Newbury House.

Schmidt, R. (1993). Consciouness, Learning, and Interlanguage Pragmatics. In: Kasper, G., Blum-Kulka, S. (Eds.) *Interlanguage Pragmatics*, pp. 21-42. Oxford: Oxford University Press.

Scott, M., (2014). *WordSmith Tools* [Computer Software]. Liverpool: Lexical Analysis Software.

Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language.* Cambridge: Cambridge University Press.

Searle, J. (1975). Indirect speech acts. In: Cole, P., Morgan, J. L. (Eds.). *Speech Acts, Syntax and Semantics, Volume 3*, pp. 59-82. New York: Academic Press.

Searle, J. (1976). A classification of illocutionary acts. *Language in Society, 5*, pp. 1-23.

Searle, J. (1979). *Expression of Meaning: Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press.

Selinker, L. (1972). Interlanguage. *IRAL 10*, pp. 209-230.

Sinclair, J. McH., Coulthard, R. M. (1975). *Towards an Analysis of Discourse: the English used by teachers and pupils*. Oxford: Oxford University Press.

Sperber, D., Wilson, D. (1995). *Relevance: Communication and Cognition*. Blackwell Publishers Ltd.

Stubbs, M. (1983). *Discourse Analysis: The Sociolinguistic Analysis of Natural Language.* Oxford: Basil Blackwell Publisher Limited.

Stubbs, M. (1996). *Text and Corpus Analysis.* Oxford: Blackwell.

The Bergen Corpus of London Teenage Language (COLT). Distributed by The University of Bergen. URL: http://clu.uni.no/icame/colt/.

The BNC Sampler, CML version. (2005). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/.

The British National Corpus (BNC), version 3 (BNC XML Edition). (2007). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/.

The Cambridge and Nottingham Corpus of Discourse in English (CANCODE), Distributed by the University of Nottingham and Cambridge University.

The Corpus of Contemporary American English (COCA). Distributed by Birgham Young University. URL: http://corpus.byu.edu/coca/.

The Diachronic Corpus of Present-Day Spoken English (DCPSE). Distributed by University College London. URL: http://www.ucl.ac.uk/english-usage/projects/dcpse/.

The Limerick Corpus of Irish English (LCIE). Distributed by The University of Limerick in conjunction of Mary Immaculate College. URL: http://www.ivacs.mic.ul.ie/corpora/.

The London-Lund Corpus of Spoken English (LLC). Distributed by University College London and Lund University. URL: http://www.helsinki.fi/varieng/CoRD/corpora/LLC/.

The Louvain Corpus of Native English Conversation (LOCNEC). Distributed by Université Catholique de Louvain.

The Louvain International Database of Spoken English Interlanguage (LINDSEI). Distributed by Université Catholique de Louvain. URL: http://www.uclouvain.be/en-cecl-lindsei.html.

Thomas, J. (1983). Cross-cultural Pragmatic Failure. *Applied Linguistics, 4*, pp. 91-112.

Thomas, J. (1995). *Meaning in Interaction: An Introduction to Pragmatics*. London: Longman.

Thornbury, S., Slade, D. (2006). *Conversation: From Description to Pedagogy*. Cambridge: Cambridge University Press.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work* (Studies in Corpus Linguistics. Volume 6). John Benjamins Publishing Company.

Tottie, G. (1991). Conversational Style in British and American English: the case of backchannels. In: Aijmer, K., Altenberg, B. (Eds.). *English Corpus Linguistics*, pp. 254-271. London: Longman.

Tottie, G. (2011). Uh and Um as sociolinguistic markers in British English. In: *The Interactional Journal of Corpus Linguistics, 16,* pp. 173-196.

University of Cambridge ESOL Examinations. (2007). *First Certificate in English Handbook for Teachers: for examinations from December 2008*. Cambridge: UCLES.

Van Dijk, T. (2008). *Discourse and Context: A Sociocognitive Approach*. Cambridge: Cambridge University Press.

Verschueren, J. (1999). *Understanding Pragmatics*. London: Hodder Arnold.

Walter, E. (2010). Using corpora to write dictionaries. In: O'Keefe, A., McCarthy, M. *The Routledge Handbook of Corpus Linguistics*, pp. 428-444. Routledge.

Weir, C. J. (2005). *Language Testing and Validation: An Evidence-Based Approach*. Basingstoke: Palgrave Macmillan.

Widdowson, H. G. (1978). *Teaching Language as Communication.* Oxford: Oxford University Press.

Wierzbicka, A. (1987). *English Speech Act Verbs: A Semantic Dictionary*. New York: Academic Press.

Wierzbicka, A. (1991). Cross-cultural pragmatics. *The semantics of human interaction.* Berlin: Mounton de Gruyter.

Wierzbicka, A. (2006). *English: Meaning and Culture.* Oxford: Oxford University Press.

Yngve, V. H. (1970). On getting a word edgewise. In*: Proceedings of the 6th Regional Meeting of the Chicago Linguistic Society*, pp. 567-577. Chicago: Chicago Linguistic Society.

APPENDIX A- Subjects' consent form

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
COMITÊ DE ÉTICA EM PESQUISA – CEP-PUCRS

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Título da pesquisa: A investigação da competência conversacional de aprendizes brasileiros de inglês em um nível intermediário de proficiência linguística.

Do convite
Por você ser aluno do curso de inglês XXX e estar no nível de proficiência linguística intermediária, você está sendo convidado (a) a participar da pesquisa: "A investigação da competência conversacional de aprendizes brasileiros de inglês em um nível intermediário de proficiência linguística".

Se decidir participar desta pesquisa, é importante que leia as informações contidas neste documento a respeito do estudo e do seu papel neste estudo. Sua participação não é obrigatória e, a qualquer momento, você pode desistir de participar e retirar o seu consentimento. Sua recusa não trará nenhum prejuízo em sua relação com os pesquisadores ou com a Pontifícia Universidade Católica do Rio Grande do Sul. É preciso entender a natureza e os riscos da sua participação e dar o seu consentimento informado por escrito ao final deste documento. Você poderá fazer todas as perguntas que precisar para entender os objetivos da pesquisa, esclarecer dúvidas acerca dos riscos, dos benefícios e outros. São-lhe garantidos esclarecimentos, antes e durante o curso da pesquisa, sobre a metodologia. Você receberá uma cópia fidedigna deste termo na qual constam as informações relativas à pesquisa bem como o telefone e endereço dos pesquisadores, por meio dos quais poderá entrar em contato para dirimir quaisquer dúvidas do projeto e de sua participação.

1. Dos pesquisadores

Esta pesquisa tem como pesquisador responsável o Prof. Dr. Jorge Campos da Costa, professor titular da Pós-Graduação em Letras da Pontifícia Universidade Católica do Rio Grande do Sul - PUCRS. Seu endereço é Av. Ipiranga, 6681. Prédio 8, sala 402. Telefone: XXX. A pesquisa conta com a co-orientação da Dr. Cristina Lopes Perna, professora colaboradora do Pós-Graduação em Letras da Pontifícia Universidade Católica do Rio Grande

do Sul – PUCRS, e com a participação da doutoranda em Linguística Cristiane Ruzicki Corsetti, professora e consultora acadêmica da XXX, local onde se dará a coleta de dados.

2. Do objetivo e da justificativa

Os temas "conversação" e "competência conversacional" são de grande relevância para professores e aprendizes de língua inglesa, visto que muitos alunos em curso de línguas objetivam comunicarem-se oralmente de forma satisfatória, mas enfrentam vários desafios durante o processo de aprendizagem. As interações verbais a que são expostos exigem habilidades específicas, além das competências gramaticais, lexicais e fonológicas. Na fala espontânea, há intenções comunicativas envolvidas que nem sempre são comunicadas explicitamente, estratégias de polidez que impactam o sucesso de mensagens comunicativas e aspectos interacionais que gerenciam uma interação verbal adequada. Porém, nem todos os aprendizes possuem conhecimento ou consciência dessas habilidades conversacionais e interacionais, pois em muitos casos, esses fenômenos não são trabalhados em sala de aula.

O objetivo deste estudo é – no lado experimental - investigar a produção oral de aprendizes adultos brasileiros no nível intermediário, enfocando aspectos conversacionais e interacionais inerentes à "competência conversacional". Objetiva-se elencar quais aspectos pragmáticos conversacionais e interacionais aparecem com mais ou menos frequência em suas produções e indicar quais aspectos parecem necessitar de maior auxílio pedagógico.

3. Dos procedimentos de coleta

A prova oral do exame de proficiência "First Certificate in English" dura 14 minutes, é realizada em duplas e constitui-se de quatro partes. Na primeira parte, o examinador faz perguntas individuais aos candidatos, de natureza pessoal. O foco é linguagem geral, interacional e social. Na segunda parte, os candidatos produzem um turno longo individual, no qual comparam duas fotos relacionadas. Os objetivos são a organização de uma unidade mais longa de discurso e a utilização das funções de comparar, descrever e expressar opiniões. Na terceira parte, os candidatos realizam uma atividade colaborativa na qual devem chegar a uma decisão conjunta. As funções objetivadas são manter uma interação, trocar idéias, expressar e justificar opiniões, concordar e discordar, fazer sugestões e especulações, avaliar e chegar a uma decisão, através da negociação. A última parte envolve uma discussão

relacionada com o tópico da parte anterior. As funções enfocadas são expressar e justificar opiniões e concordar e discordar.

Se concordar em participar deste estudo, a realização das seguintes tarefas acima lhe será solicitada. A doutoranda Cristiane Ruzicki Corsetti organizará e gerenciará entrevistas com duplas de aprendizes, que serão gravadas por um gravador de voz mp3. Após a etapa de coleta de dados, essas entrevistas serão transcritas para fins de pesquisa.

4. Dos desconfortos e riscos possíveis

A coleta de dados será realizada em salas de aula da escola XXX. As salas de aula garantem condições de trabalho seguras e tranquilas. Não há quaisquer riscos à sua integridade física ou emocional. Salienta-se, no entanto, que esta pesquisa será realizada somente se você se sentir em boas condições físicas e emocionais para realizar todas as atividades solicitadas.

5. Dos benefícios esperados

A pesquisa poderá ou não trazer-lhe benefícios com relação a tarefas de produção oral. Contudo, as informações obtidas por meio deste estudo serão relevantes para compreendermos as dificuldades interacionais e conversacionais que aprendizes de inglês no nível de proficiência intermediário apresentam durante atividades comunicativas. Além disso, os dados obtidos nesse projeto poderão servir de base a propostas metodológicas.

6. Dos custos e reembolso para o participante

Sua participação é voluntária e espontânea. Não haverá pagamento pela sua participação.

7. Da confidencialidade da pesquisa

Será garantido sigilo absoluto para assegurar a privacidade de todos os sujeitos participantes quanto aos dados confidenciais envolvidos na pesquisa. Você não será identificado quando o material de seu registro for utilizado, seja para propósitos de publicação científica ou educativa. Assim, ao assinar este consentimento informado, você autoriza as inspeções em seus registros.

8. Da Declaração de consentimento informado

Eu,.....(nome legível e por extenso), declaro que tive tempo suficiente para ler e entender as informações acima. Declaro também que fui devidamente informado pelo pesquisador Jorge Campos da Costa e pelas suas colaboradoras Cristina Lopes Perna e Cristiane Ruzicki Corsetti, sobre os procedimentos que serão utilizados, os riscos e desconfortos, os benefícios, o custo/reembolso, dos participantes e a confidencialidade da pesquisa. Confirmo que toda a linguagem técnica utilizada na descrição da pesquisa foi satisfatoriamente explicada e que recebi respostas para todas as minhas dúvidas. Declaro ainda que me foi assegurado que posso retirar o consentimento a qualquer momento, sem que isso leve a qualquer penalidade ou a perda de benefícios. Confirmo ainda que recebi uma cópia desse Termo de Consentimento informado.

Caso tiver novas perguntas sobre esse estudo, posso chamar o Prof. Dr. Jorge Campos da Costa, a Prof. Dr. Cristina Lopes Perna e a doutoranda Cristiane Ruzicki Corsetti no seguinte endereço: Av Ipiranga, 6681. Prédio 8, sala 402. Telefone: XXX ou o Comitê de Ética de Pesquisa da PUCRS: 51-33203345.

Dou meu consentimento de espontânea vontade e sem reservas para participar deste estudo.

Assinatura do (a) participante: _____

Data: ____/_____/2012

Eu.....

atesto que expliquei cuidadosamente a natureza e o objetivo deste estudo, os possíveis riscos e benefícios da participação nesta pesquisa. Acredito que o (a) participante recebeu todas as informações necessárias, as quais foram fornecidas em uma linguagem adequada e compreensível, e que o (a) participante compreendeu tais explicações.

Assinatura do pesquisador: _____

Data: ____/____/2012

ATTACHMENT A- FCE speaking test public version

First Certificate in English: Handbook for Teachers (UCLES, 2007, p.79-81)
Part 1: 8 minutes

Introduction

Good evening. My name is …. And your names are (subject A and subject B)?
First of all, I'd like to know something about you.

- Where are you from (A)?
- And you (B)?
- What do you like about living in Porto Alegre (B)?
- And what about you (A)?

Topic: Likes and dislikes

- **Do you prefer to spend time on your own or with other people (A)? (Why?)**
- **What's your favourite food (B)? (Why do you like it?)**
- **Tell us about a day you've really enjoyed recently (A).**
- **Do you like going to parties (B)?**

Topic: Free time

- **How much time do you spend at home (A)? (What do you enjoy doing?)**
- **Who do you spend your free time with (B)? (What sort of things do you do together?)**
- **Have you got any plans for this weekend (A)? (What are you going to do?)**
- **Does anyone you know have an interesting hobby (B)? (What does he/she do?)**

Topic: Holidays and Travel

- **Which area of your country would you like to get to know better (A)? (Why?)**
- **What's the most interesting place you've visited near here (B)? (Tell me about it)**
- **Do you plan your holidays carefully or do you prefer to just go (A)? (Why)**
- **Where would you like to go on holiday in the future (B)? (Why)**

Topic: Media

- **How much TV do you watch in a week (A)? (Would you prefer to watch more TV than that or less?)(Why?)**
- **Tell us about about a TV programme you've seen recently (B).**
- **Do you have a favourite newspaper or magazine (A)? (Why do you like it?)**
- **Do you use the Internet to learn new things (B)? (What sort of things do you look for?)**

Part 2: 4 minutes

In this part of the test, I'm going to give each of you two photographs. I'd like you to talk about your photographs on your own for about a minute and also to answer a short question about your partner's photographs.

(Subject A), it's your turn first. Here are your photographs. They show **people making music in different ways**.

(Place photos in front of A)

I'd like you to compare the photographs, and say **why you think the music is important to the different groups of people**.

All right?

(Subject A) 1 minute

Thank you. (Subject B), **which type of music would you prefer to listen to?**
(approximately 20 seconds)

Now (B), here are your photographs. They show **people of different ages on educational visits**.

(Place photos in front of B)

I'd like you to compare the photographs, and say **what you think the people will learn on their visits**.

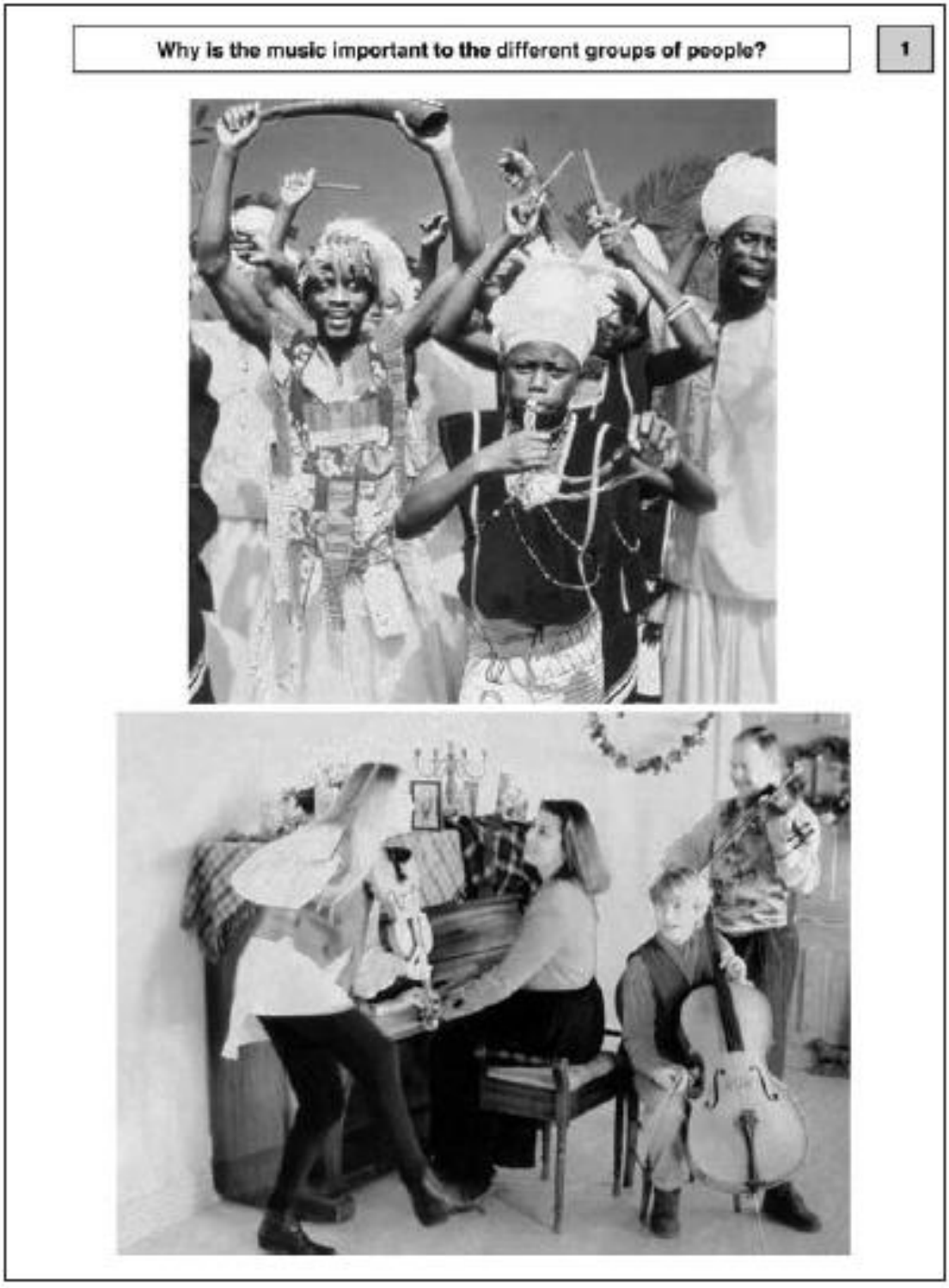
All right?

(Subject B) 1 minute

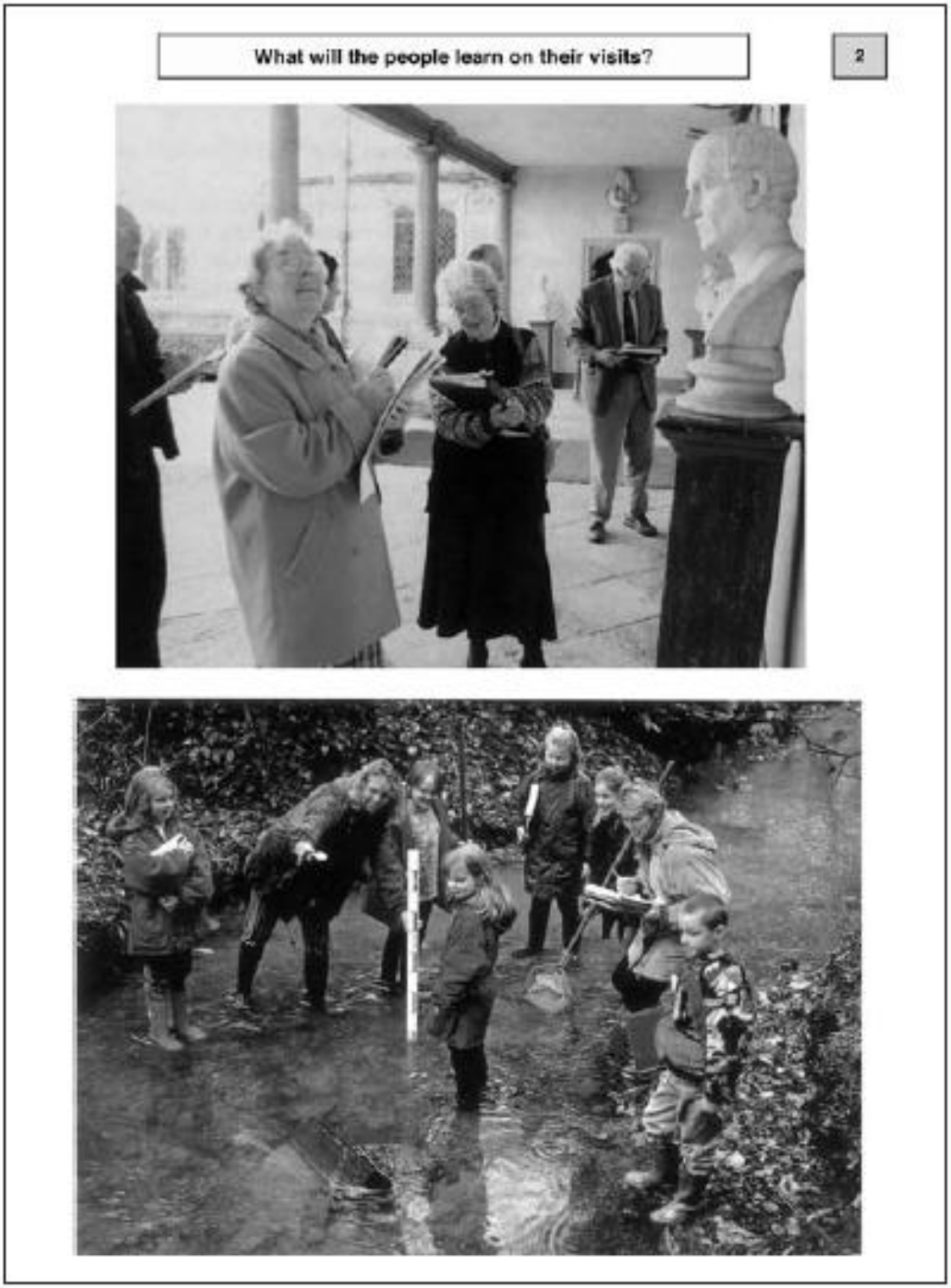
Thank you. (Subject A), **which of these things would you like to learn about?**
(approximately 20 seconds)

Thank you.

Part 2: Subject A's photographs



Part 2: Subject B's photographs



Parts 3 and 4: 7 minutes

Part 3:
Now I'd like you to talk about something together for about three minutes.
I'd like you to imagine that a local café wants to attract more people. Here are some suggestions they are considering.
(Place visual material in front of subjects)

First talk to each other about **how successful these suggestions might be**. **Then decide which two would attract most people**.

All right?

(Subjects A and B) 3 minutes

Part 4:

Select any of the following questions as appropriate. Subjects A and B discuss questions together.

- **Would you like to spend time in a café like this?... (Why?/Why not?)**
- **Would you like to work in a café?... (Why?/Why not?)**
- **What sort of restaurants are most popular with visitors in your country?... (Why?)**
- **What sort of things do people complain in restaurants?**
- **Young people usually go to different places to relax than older people. Why do you think this is?**
- **Some people say that going out to relax is a waste of time and money. Do you agree?... (Why?/Why not?)**

Thank you. That's the end of the interview.

Part 3: Visual material