VERÔNICA DA SILVEIRA PEDRO

CONNECTED DISCOURSE AND THE PHRASEOLOGY OF SPOKEN ENGLISH: A CORPUS STUDY OF AFFIRMATIVE RESPONSES

Tese apresentada ao Curso de Pós-graduação em Letras da Universidade Federal Fluminense, como requisito parcial para obtenção do Grau de Doutora. Área de concentração: Estudos Lingüísticos

Orientadora: Profª Drª SOLANGE COELHO VEREZA

Niterói
2006

Ficha Catalográfica elaborada pela Biblioteca Central do Gragoatá

P372 Pedro, Verônica da Silveira.
Connected discourse and the phraseology of spoken English: a corpus study of affirmative responses / Verônica da Silveira Pedro. – 2006.
164 f.
Orientador: Solange Coelho Vereza.
Tese (Doutorado) – Universidade Federal Fluminense, Instituto de Letras, 2006.
Bibliografia: f. 143-151.

1. Língua inglesa – Gramática. 2. Língua inglesa – Lingüística. 3. Língua inglesa – Lexicografia. I. Vereza, Solange Coelho. II. Universidade Federal Fluminense. Instituto de Letras. III. Título.

CDD 425

BANCA EXAMINADORA

Profª Drª SOLANGE COELHO VEREZA – Orientadora
UFF

Profª Drª MARIA CARMELITA PÁDUA DIAS
PUC-Rio

Profª Drª HELENA FRANCO MARTINS
PUC-Rio

Profª Drª REJANE TEIXEIRA VIDAL
UFF

Profª Drª TEREZINHA BENEVIDES LOBIANCO
UFF

Prof. Dr. DAVID SHEPHERD (suplente)
UFF

VERÔNICA DA SILVEIRA PEDRO

CONNECTED DISCOURSE AND THE PHRASEOLOGY OF SPOKEN ENGLISH: A CORPUS STUDY OF AFFIRMATIVE RESPONSES

Tese apresentada ao Curso de Pós-graduação em Letras da Universidade Federal Fluminense, como requisito parcial para obtenção do Grau de Doutora. Área de concentração: Estudos Lingüísticos

Aprovada em agosto de 2006.

BANCA EXAMINADORA

_____

Profª Drª SOLANGE COELHO VEREZA – Orientadora
UFF

_____

Profª Drª MARIA CARMELITA PÁDUA DIAS
PUC-Rio

_____

Profª Drª HELENA FRANCO MARTINS
PUC-Rio

_____

Profª Drª REJANE TEIXEIRA VIDAL
UFF

_____

Profª Drª TEREZINHA BENEVIDES LOBIANCO
UFF

Niterói
2006

AGRADECIMENTOS

À Profª Drª Solange Coelho Vereza pelo constante incentivo, dedicação, indicações bibliográficas e apoio durante toda a realização desta pesquisa,
À professora Drª Rejane Teixeira Vidal,
À minha família.

RESUMO

Este trabalho apresenta uma investigação das respostas afirmativas contidas em um corpus de inglês falado língua materna. A fundamentação teórica é baseada em estudos nas áreas de pragmática, colocação, fraseologia, lingüística de corpus e gramática de padrões (*pattern grammar*). Os pressupostos teóricos foram apresentados e discutidos. As respostas afirmativas dos tipos *yeah*-responses e *yes*-responses contidas no corpus e as iniciativas que as originaram foram identificadas em seus valores ilocucionários e quantificadas. Os valores ilocucionários mais freqüentes da iniciativas e respostas foram comparados e contrastados em relação às suas ocorrências nos dois tipos de respostas. As respostas cujos valores ilocucionários foram mais freqüentes foram, então submetidas a análises mais profundas. Primeiramente, foram identificadas aquelas que possuíam enunciados proferidos em continuação às palavras *yeah* e *yes*. Este grupo de respostas foi analisado, inicialmente, em relação às diferentes formas dos enunciados nelas contidos. Foram classificadas em respostas "independentes" e "complexas". As respostas independentes eram constituídas por "expressões independentes" e as "complexas" eram constituídas por sequências de expressões (em enunciados), as quais incluíram, em alguns casos, o uso das "expressões independentes". Os padrões de ocorrência das expressões independentes foram identificados. Em seguida, foram identificadas as colocações das palavras *yeah* e *yes* nas respostas que continham continuações às mesmas, e foram identificados os seus padrões de ocorrência. As respostas com o valor ilocucionário mais freqüente e que continham continuações, identificadas no corpus como "confirmações", foram selecionadas para as análises dos padrões de ocorrência de formas e tempos verbais, dos elementos de coesão e das expressões lexicais recorrentes. Os padrões recorrentes de uso destes elementos foram identificados e quantificados. Verificou-se que as respostas afirmativas que contêm *yeah* ou *yes* no corpus analisado ocorreram na fala em padrões recorrentes em relação às iniciativas, especialmente as mais freqüentes; as colocações de *yeah* e *yes* também ocorreram em padrões observáveis, assim como as expressões independentes e as respostas complexas. Nestas últimas, verificou-se que as continuações à palavra *yeah* (escolhida em um recorte de pesquisa) continham verbos, elementos coesivos e sequências de expressões lexicais que ocorreram no corpus em padrões de colocação e freqüência.

Palavras-chave: lingüística de corpus; colocações; expressões lexicais; fraseologia; gramática de padrões; respostas afirmativas em inglês língua materna.

ABSTRACT

The present study presents an investigation of affirmative responses in a corpus of spoken English as a mother tongue. The theoretical background consists of a discussion of issues in spoken English and connected spoken discourse in light of the contributions from the fields of pragmatics, corpus linguistics, studies into collocation, phraseology and pattern grammar. Some relevant aspects of the literature in these fields are presented and discussed. The *yeah* and *yes* responses encountered in the corpus and their respective initiation moves were analysed in relation to their illocutionary values. The latter were identified and quantified. The most frequent illocutionary values of both the initiation moves and the responses were compared and contrasted. The responses which belonged to the most frequent illocutionary value were then submitted to deeper analyses. Firstly, the responses which contained continuations to the words *yeah* and yes *were* identified and quantified. Secondly, they were classified into independent and complex responses. Independent responses are those which contain independent phrases, whereas complex responses consist of responses which contain sequences of phrases, including, in some cases, instances of independent phrases. The patterns of occurrence of these responses were then identified. The words *yeah* and *yes* were also investigated in relation to their collocates within those responses which contained continuations. The responses which had the most frequent illocutionary value, namely, confirmings, were analysed in relation to the occurrence of verb forms and tenses, cohesive items and lexical phrases. Their recurrent patterns of use were identified and quantified. The analyses thus demonstrated that the *yeah* and *yes* responses encountered in the corpus of spoken English occurred in recurrent patterns of combinations of initiation moves and responses, collocations of word-forms and phrases, and independent and complex responses. In the latter case, the continuations within the responses contained verbs, cohesive items and sequences of lexical phrases which occurred in the corpus in patterns of frequency and sequences of phrases.

Key-words: corpus linguistics; collocation; lexical phrases; phraseology; pattern grammar; affirmative responses in English as a mother tongue.

LIST OF ABBREVIATIONS

CP	Cooperative principle
EFL	English as a foreign language
ELT	English language teaching
EL1	English as a mother tongue
EL2	English as a second language
EFL	English as a foreign language
FTA	Face-threatening act
IF	Illocutionary force
IFID	Illocutionary force indicating device
IR	Independent responses
IM	Initiation move
Ls	Learners
L1	First language
L2	Second language
LP	Lexical phrase
NS	Native speaker
NNS	Non-native speaker
S1	Speaker 1 (who utters an initiation move)
S2	Speaker 2 (who utters a response)
Ss	Speakers
T	Turn in conversation
YC	*Yes* corpus
YhC	*Yeah* corpus
YhR	*Yeah*-response
YsR	*Yes*-response
YYR	*Yeah* and *yes* responses

CONTENTS

1- INTRODUCTION

1.1- The importance of research into connected discourse

This study aims at presenting an analysis of connected discourse in English as a mother tongue through an investigation of a corpus of affirmative responses[1] containing the words *yeah* and *yes* as used in real-life spoken interaction. The major aims of the analysis are the identification of the recurrent pragmalinguistic patterns and roles of *yeah* and *yes* responses (addressed in the initial chapters as YYRs) and their continuations (*i.e.* the utterances that follow the words *yeah* and *yes* as uttered by speaker 2[2]; *cf*. Chapter 5.1) in the connecting of discourse. Therefore, all the constituents of the YYRs obtained from a corpus[3] of spoken English as a mother tongue which were uttered after an initiation move[4] by a speaker's interlocutor have been analysed. These constituents include the words *yeah* and *yes* and all the utterances prior or subsequent to them which were produced by the same speaker. In addition to that, the initiation moves that generated them have also been analysed[5].

The YYRs under analysis are responses which contain the words *yeah* or *yes* alone, at initial, medial or final position. The YYRs which contain *yeah* or *yes* alone were included in the initial analyses (for quantification purposes) but were not the focus of the major analyses developed. The YYRs were collected from the *Bank of English* corpus of spoken English, which is developed by the COBUILD[6] Project at the University of Birmingham. All instances

---

[1] We have decided to use the term "response" as a generic term of reference to the words *yeah* and *yes* which are cited in the present study, since these words were used as responses in over 99% of cases in the corpus.

[2] Speaker 2 (S2) is the person who has the turn in conversation and begins to speak after being addressed by another speaker who previously had the turn. In the present study, S2 is, generally, the person who utters a response in the corpus.

[3] A corpus is defined by Sinclair (1991, p. 171) as "a collection of naturally-occurring language text, chosen to characterise a state or variety of a language."
Hunston (2002, p. 2) notes that "More recently, the word has been reserved for collections of texts (or parts of text) that are stored and accessed electronically. […] A corpus is planned, though chance may play a part in in the text collection, and it is designed for some linguistic purpose. The specific purpose of the design determines the selection of texts, and the aim is other than to preserve the text themselves because they have intrinsic value. […] The corpus is stored in such a way that it can be studied non-linearly, and both quantitatively and qualitatively."

[4] A move is defined by Richards (1992, p. 238) as "a unit of discourse which may be smaller than an utterance."

[5] The choice has been made for a corpus of spoken English as a mother tongue due to the fact that the present research is directed to an audience who are interested in studies in English as a mother tongue.

[6] COBUILD is an acronym for Collins Birmingham University International Database.

of *yeah* and *yes* in the spoken corpus were analysed and a total of one thousand nine hundred and seventy-two utterances (including the initiation moves and the responses) were encountered.

The importance of studies into connected spoken language is well described by Stubbs (1983). In a brief description of some of the issues which can be encountered in linguistic studies, the author mentions that studies in Linguistics have been "[…] based on a restricted range of linguistic data. The data have often been simplified, throughout the last two thousand years of grammatical study, by being drawn from written language or from introspection […]" (STUBBS 1983, p. 82). He notes that "[…] since conversational language has not been systematically described, it is in any case easier to focus attention on permanently recorded written languages […]" (*ibid.*). In addition to that, according to him, "[…] introspective judgements have themselves been unduly influenced by written language." (*ibid.*).

Stubbs reinforces his view about the importance of research which focuses on natural connected discourse by arguing that extra-sentential phenomena are important to understanding intrasentential elements, and that, therefore, "many details of linguistic organisation will be missed, if language in use in connected discourse is ignored" (STUBBS, 1986, p. 82). The author claims that a broader pragmatic perspective would be necessary to the understanding of the "syntagmatic chaining of clauses and sentences in larger exchanges or sequences" as "certain phenomena involving particles, adverbs and conjunctions can only be explained with reference to the syntagmatic chaining of linguistic units at the clause or above" (*ibid.*, p. 83).

Another argument in favour of investigations into the use of affirmative responses (of any type) in connected natural discourse is the fact that a large amount of the recent studies in pragmatics and speech act theory have often focused on speech acts that convey illocutionary values of conflict, such as denial, refusal, request, complaint and disagreement, among others.

However, the notion of spoken discourse as a matter of keeping social bonds also leads us to the fact that speakers aim, most of the time, at the maintenance of these bonds and the saving of alter's face[7] in discourse. Therefore, the negotiation of meaning is a characteristically permanent feature of discourse, be it in situations of conflict or not. It requires speakers' permanent use of skills and strategies for the maintenance of balance during the negotiation process. However, most research on pragmatics does not, generally, account for the linguistic features employed in such circumstances and a study of the use of

[7] The notion of *face* relates to 'an individual's self-esteem' in the context of discourse, according to Brown and Levinson (1978 [1987], p. 2).

affirmative responses in actual discourse can play an important role in bringing to light some of the various pragmalinguistic features involved.

The case for investigations into the pragmalinguistic properties of responses in general is an interesting one, especially with a view to identifying their roles in structuring connected, chain-style spoken discourse. As Stubbs (*ibid*.) pointed out

> connected discourse is clearly not random. People are quite able to distinguish between a random list of sentences and a coherent text […] (*ibid*., p.19).
> […] To say that discourse has structure does not necessarily mean that this structure is specifically linguistic […]. The structure may be the surface manifestation of much more general organization, including the causal relation between events in the world and our inferences about such events. [*Ibid*., p. 102, 103]
> […] it therefore makes sense to study the relationship between units above the clause, possibly reclassifying these units as functional acts or moves. [*Ibid*., p.104]

The present research, with its focus on affirmative responses containing *yeah* or *yes*, represents an attempt at the identification of the underlying organizational features of such responses since, as Stubbs (1983) has pointed out,

> *Yes* and *no* are oddities on syntactic grounds. They are often regarded as clause or sentence substitutes. However, as Halliday and Hasan (1976:137-8) point out, they are better regarded as elliptic forms, since they express just the polarity of the clause. […] *Yes* and *no* can also co-occur with other elliptic forms of the clause. (STUBBS, *ibid*., p. 111.)

Still according to the author, both *yes* and *no* are fundamentally an interactive or discourse phenomenon, since they are responses to preceding utterances. Although they are normally regarded as "a matter of syntax or even lexis" (*ibid*., p. 111), they can only be understood in their contexts of occurrence, which would provide "a motivation for going beyond lexico-syntax" (*ibid*.).

1.2 – Traditional approaches to responses in ELT: a problem area

Through my own teaching experience, I have identified a problem area in the oral production of affirmative responses by Brazilian learners of English as a foreign language (EFL), as well as in the fields of theoretical and practical work in Applied Linguistics (*cf*. SILVEIRA PEDRO, 1999). Brazilian learners (and perhaps learners of English as a foreign language of other nationalities too) have frequently faced difficulties in the production and conveyance of meaning in spoken discourse. Although the recent literature in acquisition and interlanguage pragmatics has identified various factors that may affect learners' production of discourse (see, for example, ELLIS 1994, Chapters 2, 3 and 4 for discussions of the issues

involving social factors, language transfer and cognition), if we take a closer look at one of the most basic forms of carrying the spoken discourse further[8], *i.e.*, through the utterance of affirmative responses in conversation, we are faced with the fact that this is an area that can present difficulties to non-native speakers of English and which has been neglected in the literature. Since such difficulties can be frequently identified in learners' production, we can argue that they affect aspects of continuity and meaning negotiation in communication. In addition to that, these difficulties have also been neglected by writers of materials for the teaching of English as a foreign language.

Some explanations to these difficulties might include the following issues: firstly, it could be argued that native speakers and Brazilian learners of English as a foreign language make use of different linguistic structures when responding affirmatively: while *yes*-responses appear in English at a very high frequency, their equivalent form in Portuguese (responses containing the word *sim*) occurs only occasionally. The Portuguese equivalent *sim* is rarely used in everyday dialogues in affirmative responses meaning acceptance, agreement, confirmation and others, its use being more frequent as a conversation marker of acknowledgement in more formal contexts or in contexts in which spontaneous cooperation does not occur naturally. On the other hand, other forms are more frequent in natural everyday conversations. For example, in Portuguese the preferred sequences may include the use of expressions such as *Hã, hã,* a sentential subject followed by repetition of the verb in the affirmative form (*Ele é, Eu vou, Ele sabe*), or an elliptical subject with an affirmative verb (*É, Vou, Sabe*). Therefore, and not surprisingly, aspects of first language transfer into the target language, among others, are likely to play an important role in generating pragmalinguistic failure[9]. An example of the strategies used by adult Brazilian students of English (at beginner's or elementary levels) is the use of their mother tongue in an attempt to fully express what they intend to communicate, since they perceive that the forms that are usually presented in coursebooks do not entirely convey the intended meanings.

Secondly, as already mentioned in the previous paragraph, during the learning process learners of English as a foreign language are taught how to answer questions affirmatively through the use of *yes*-fronted short-answers in most instances. These sentence forms are the

[8] 'Carrying the discourse further' is used here in parallel to what McCarthy (1991, p. 70) calls 'the forward moving of the discourse'. He notes that 'utterances by one speaker are an invitation to a response by another […]; the initiating utterance puts an obligation on the responding speaker to make his/her turn both relevant […] and a positive contribution to the forward moving of the discourse […]'.

[9] Thomas (1983, *apud* ELLIS, 1994, p. 165) notes that *pragmalinguistic failure* "occurs when a learner tries to perform the right speech act but uses the wrong linguistic means (*i.e.*, deviates with regard to appropriateness of form)", whereas *sociopragmatic failure* "takes place when a learner fails to perform the illocutionary act required by the situation (*i.e.*, deviates with regard to appropriateness of meaning)."

most frequent ones presented to learners in classroom materials at the initial levels, regardless of their appropriateness to context. Some examples include

1- the presentation of verb tenses in the affirmative, negative and interrogative forms, in which short answers are presented as the possible structure for the realisation of answers to questions in most cases. The questions and answers are often presented in lists of sentences to be completed, with no reference to context and no special emphasis on the answers. Some examples include: (from an exercise in which students are asked to complete the sentences after the presentation stage for the present simple) "____ she drink coffee? – Yes, she ____."; "____ he live in Paris?" – No, he ____."; (from an exercise in which students are asked to "make present perfect sentences") "/you ever been/Japan? Yes, I /."
(from the coursebook *English File 1* [10]);

2- one case of affirmative responses different from short answers was found in English File 1 (*ibid.*). They consisted of the phrases "I think so", "I hope so" and "Maybe", which were meant for use after "going to" questions. The phrases, however, did not contain "*yes*" at initial position nor any type of continuation. It should be noted here, that I do not mean that "*yes*" is essential to the structure of the response but the contradiction lies in the fact that there was one short answer on the same list ("*Yes, I am*"), thus imparting a type of divide between their uses. Furthermore, the fact that no continuations were presented aggravated the lack of communication in an exercise that could well have served for an increase in learners' communicative spoken abilities;

3- one single exercise in the coursebook *Cutting edge: pre-intermediate* [11] presented short dialogues for gap filling in some questions and answers. The affirmative answers contained continuations and were good examples of the type of responses found in natural conversation. However, after completing the gaps, students were asked to listen to the tape and check their answers. Then they had to identify the speakers who were making requests and those who were asking for permission. In other words, the exercise did not aim at practising spoken English nor did it aim at acquiring any types of responses.

These three examples demonstrate that the various other possible realisations of affirmative responses appear nearly always as accessories to other functions being taught.

[10] OXENDEN and SELIGSON, 1996.
[11] CUNNINGHAM and MOOR, 2001.

Thirdly, learners are systematically taught how to answer questions whereas there seems to be no systematic teaching of how to respond to utterances of various illocutionary values[12] that belong to other sentence types (such as requesting and seeking confirmation, giving advice, and others which can be realised through the use of statements or commands). For instance, in my previous research on this subject I demonstrated that responses realized through statements represent native speakers' preferred strategy for the elicitation of responses in spoken English, rather than questions (*cf*. SILVEIRA PEDRO, 1999).

In addition to that, since the various possibilities for the realisation of affirmative responses are not systematically presented to learners, a deriving problem arises: the connections between context and discourse are seldom made explicit to them, which increases the chances of pragmalinguistic failure.

Another aspect seems to be the age component as, according to my own teaching experience, different age-groups react differently to such deficiencies. Children seem to adhere instantly to the traditional forms presented in textbooks; adolescents only partially do so; adults, however, rarely conform to merely uttering *Yes, I do/did/can*/etc. The latter group frequently tries to express other meanings (in other words, they try to 'further' their discourse) and immediately makes use of transfer from L1, translation or questions about how to say what they want to communicate in English. The whole process usually takes place without learners' conscious criticism of the contents of coursebooks; however, sometimes they express some dissatisfaction with the limitations of short-answers. The individual, therefore, finds little room for expressing him/herself in the discourse, being unable to exploit the different nuances of meaning. In summary, transfer from the mother tongue and the lack of knowledge about the contextual constraints involved in each situation seem to play a decisive role in instances of pragmalinguistic failure at various levels in the production of discourse.

This demonstrates that the way responses are approached in coursebooks (which may include various speech acts a speaker may perform when responding to an interlocutor, be it affirmatively or negatively), as in the case of YYRs, often neglects some very frequent pragmalinguistic features. This, therefore, may affect learners' potential for carrying the discourse further or for producing more natural, connected discourse.

Table 1 illustrates some of the contrasts between the production of *yes*-responses by native speakers and the forms commonly found in coursebooks the teaching of English (*cf*. the

[12] The illocutionary value of an utterance is the communicative force of an utterance or the "type of function of language" (AUSTIN, 1962, p.100); *cf.* the discussions of pragmatics and speech act theory in Chapter 2 of the present study.

materials discussed in the preceding paragraphs) as a foreign language. The information provided in table 1 derives from analyses I carried out prior to the present research: in the case of the production by native speakers, the information presented consists of data I collected and analysed in a pilot study into affirmative responses, and, in the case of coursebooks for the teaching of EFL, from an analysis of the language input they provide.

Table 1: Comparative table of the different uses of *yes*-responses in the discourse of native speakers and in the structures commonly found in EFL coursebooks

Naturally-occurring *yes* responses in discourse	***Yes* responses in the discourse of EFL coursebooks**
1. Multiple discourse structures.	1. A limited number of discourse structures is presented (usually grammatical structures of the short-answer type).
2. Communicative value/meaning is totally conveyed.	2. Meaning is partially conveyed, usually.
3. Contextualisation; context is usually known by participants.	3. Partial or inadequate contextualisation; background information is little known.
4. Pragmalinguistic realisation related to context.	4. Frequent pragmalinguistic failure may occur in learner output, as a consequence of input.
5. Responses are used in the negotiation of meaning/link parts of the discourse.	5. Discourse is often interrupted after short-answers.
6. Responses often include some type of extension that upgrades or mitigates their meanings.	6. *Yes*-responses rarely include continuations or extensions.
7. The different contexts originate various pragmalinguistic solutions.	7. The same structures are presented independently of context, usually.
8. *Yes*-responses can be used as politeness strategies.	8. Their use as politeness strategies is not made explicit to learners. This may cause the opposite effect of bluntness (through the use of short-answers).
9. Intonation varies according to context.	9. Intonation is usually little exploited.
10. Short-answers are rarely used.	10. Short-answers are very frequent.

(*Cf*. SILVEIRA PEDRO, 1999)

These insights reinforce the need for systematic investigations into the multifunctional nature of *yes*-responses, their different realisation patterns and their contribution to meaning negotiation. An awareness of their constituents is fundamental to a better understanding of their regularities in language.

1.3 – The importance of research into phraseology

In my Master's dissertation I developed a study which focused on a smaller corpus of *yes*-responses (which did not include all the *yes* responses used in the present study nor included any instances of *yeah* responses). Those investigations departed from a classification of the utterances under analysis into sentence types, thus constituting a totally different type of research. However, I identified the occurrence of word combinations that appeared at high frequencies and were best described as *lexical phrases* (*cf*. NATTINGER and De CARRICO, 1992), which can be initially defined as

> 'chunks' of language of varying length […] multi-word lexical phenomena that exist between the traditional poles of lexicon and syntax, conventionalized form/function composites that occur more frequently and have more idiomatically determined meaning than language that is put together each time […] These phrases include short, relatively fixed phrases […], or longer phrases or clauses […] each with a fixed, basic frame, with slots for various fillers[…] Each is associated with a particular discourse function […]. (NATTINGER and De CARRICO, 1992, p. 1)

Nattinger and DeCarrico (*ibid.*) differentiate lexical phrases from collocations and ordinary syntactic strings that do not count as lexical phrases. They describe them as collocations that have been assigned pragmatic functions.

The authors propose two types of lexical phrases. The first would be "strings of specific (non-productive) lexical items, which allow no paradigmatic or syntagmatic substitution" (NATTINGER and De CARRICO, *ibid.*, p. 36), and the second, generalized "productive frames" which consist of "strings of category symbols (or otherwise generally specified syntactic/semantic features) and specific lexical items, which have been assigned a pragmatic function" (*ibid*.). They exemplify the former type with the expressions *"what on earth*", *"at any rate"*, *"by and large"* and *"as it were"*, and of the latter with *"a* + N [+time] + *ago*", and "Modal + *you* + VP". These frames would "underlie specific lexical phrases, such as *"a year ago, a month ago,* and *would you pass the salt*?, *could you shut the window*?, etc." (*ibid.*, p. 37).

Other examples of lexical phrases include *"I think (that) X*", *"my point is that X*", *"let me start by X*", *"it seems (to me) that X*" (*ibid.*, p. 43-44). These will be further discussed in Chapters 4 and 6.

Therefore, the findings in my previous study revealed that lexical phrases are recurring features of *yes*-responses uttered in connected discourse, thus drawing attention to the need for further investigations into the field of phraseology. The present research aims at contributing to such studies, particularly to an understanding of phraseology and its relationship with pragmatics. Within the latter, it could be hypothesised that the politeness principle (*cf*. Chapter 2) performs an important role in the case of affirmative responses since it can influence the choice of the linguistic realisations of lexical phrases and other constituents available in the language.

Studies into pragmatics can benefit from a focus on phraseology. This can be exemplified by Moon's (1997) discussions of "multi-word items". The author proposes that "the examination of texts shows up the crucial importance of this: the multi-word items chosen are not arbitrary or causal, but integral parts of the whole discourse." (MOON, 1997, *apud* SCHMITT and McCARTHY, 1997, p.56-57.)

Moon's (*ibid*.) observations about the choice of multi-word items in natural discourse are noteworthy, since they reinforce the argument developed in the previous paragraphs: multi-word items, as well as other items of connected discourse, when examined in actual uses in texts, reveal that their choices "are not arbitrary or causal, but integral parts of the whole discourse" (*ibid*). Therefore, more pragmatic approaches to these items would cast a new light on aspects of language use that are usually given less prominence in linguistic studies.

1.4 - Research aims and research questions

The general aim of the present study is, thus, to present an analysis of connected discourse in English as a mother tongue through an investigation of a corpus of affirmative responses containing the words *yeah* and *yes* as used in real-life interaction.

The more specific aims are the following:

1- To identify any existing patterns of illocutionary values of *yeah* and *yes* responses and their initiation moves in a corpus of spoken English.

2- To investigate the collocational patterns of *yeah* and *yes* in clause and phrase responses, to compare and contrast them and discover if they are identical or different.

3- To identify any patterns of associations between grammar and lexis (*e.g*., the lexical classes of verbs, adjectives, and verbs, in negative and affirmative forms, that may co-occur in YYRs).

4- To identify the most frequent patterns of cohesion.

5- To identify any frequent patterns of uses of lexical phrases in YYRs and their role in connecting discourse.

These aims are in consonance with the following research questions and hypotheses:

Specific research questions and hypotheses:

Question 1

Do the words *yeah* and *yes* have any patterns of prevailing illocutionary values in a corpus of spoken English? If they do, are they the same for both words? Are their any patterns between *yeah* and *yes* responses and the initiation moves that originate them?

Hypothesis 1

The words *yeah* and *yes* have patterns of prevailing illocutionary values in a corpus of spoken English. These patterns are not identical. There are patterns of occurrence between the illocutionary values of *yeah* and *yes* responses and their initiation moves.

Question 2

Do the words *yeah* and *yes* when used in clause or phrase responses have any patterns of collocation? If they do, are the patterns the same or different?

Hypothesis 2

Yeah and *yes*, when used in clause or phrase responses, like many words, co-occur with other words. They have collocational patterns which are not identical.

Question 3

Do *yeah* and *yes* have any lexical selection preferences? If positive, are these preferences associated with any particular syntactic structures?

Hypothesis 3

Yeah and yes *have* lexical selection preferences and these are associated with particular syntactic structures.

Question 4

Are there any observable patterns of cohesion in YYRs?

Hypothesis 4

YYRs contain items of cohesion which occur in patterns in most cases.

Question 5

Do *yeah* and *yes* responses occur in any patterns of lexical phrases?

Hypothesis 5

Yeah and *yes* responses occur in patterns of lexical phrases within a speaker's turn[13].

The next chapter will present an overview of the contributions of pragmatics and speech-act theory to studies into natural spoken discourse. We shall present some working definitions of pragmatics and will address pragmatics and speech-act theory from a historical and critical perspective. The contributions of the leading researchers in the field will be briefly discussed. Furthermore, we shall present some considerations about some pragmatic factors which are involved in the production of affirmative responses and conversational sequences.

Chapter 3 will discuss the contributions of corpus linguistics to studies to phraseology, with a view to presenting a better understanding of some phenomena that underlie the organization of connected spoken discourse. It includes insights into the focus on naturally-occurring discourse, which has greatly benefited from the contributions of corpus linguistics and the use of computerised corpora. The importance of the evidence obtained through the use of the computer is emphasised, especially with respect to the contributions to studies into collocation and/or phraseology. These studies are also discussed from the perspective of two principles of the organization of language, namely, the idiom principle and the open-choice principle, which were formulated in a parallel with the development of corpus studies.

Another outcome of corpus studies that will be addressed in Chapter 3 is the notion that very frequent words (and their most frequent senses) are useful features of the language. This notion has permeated most of the theoretical considerations in the field. Consequently, researchers have endeavoured in re-thinking the boundaries between grammar and lexis, and a discussion of this new paradigm is included in the chapter.

[13] A turn in conversation is thus described by Richards (1992, p. 390): "The person who speaks first becomes a listener as soon as the person addressed takes his or her turn in the conversation by beginning to speak."

Chapter 4 will present a discussion of the importance of studies into phraseology for a better understanding of the organisation and structuring of discourse. The chapter will address the notions of predictability in discourse and phraseological competence, which account for various linguistic phenomena. Furthermore, the current terminology and typology which are frequently encountered in the literature will be explored.

Chapter 5 will describe the methodology, which includes the use of the Bank of English corpus of spoken language, the identification and classification of the *yeah* and *yes* responses and their initiation moves into pragmatic categories, the identification of the most frequent initiation moves, *yeah* and *yes responses*, the different types of responses, the collocations of *yeah* and *yes*, the analysis of the most frequent responses (*i.e., confirming*) and the description of the scheme for analysis.

Chapter 6 will present the analyses of the initiation moves and *yeah* and *yes* responses in the corpus. The initial findings will be compared and contrasted. Following that, the general distribution of the responses which contained continuations will be analysed and the most frequent type of response with continuations will be analysed in relation to its collocational and phraseological characteristics. The investigations into the phraseology of *yeah* and *yes* responses will include the analyses of the collocations of *yeah* and *yes*, the patterns of lexical phrases in the most frequent responses, and the patterns of the verbs and items of cohesion that the latter contain. Finally, the results obtained from the analyses will be discussed.

Chapter 7 will present some conclusive remarks about the analyses and the theoretical background, thus offering some the insights into the field.

2- SPEECH ACT THEORY AND THE UNDERSTANDING OF MEANING

2.1- Introduction

The theory of speech acts is believed to have its roots in different schools of thought throughout history. Kerbrat-Orecchioni (2005) argues that different fields of knowledge from the past had explored the idea that "saying is also doing" (KERBRAT-ORECCHIONI 2005, p.17; translation mine). The author mentions the case of rhetoricians, whose focus of investigations was placed on the art of persuasion through the use of discourse. She also highlights the fact that in the Middle Ages certain "performative conceptions of the activity of language" (*ibid*.) were developed. However, it was in the twentieth century that an awareness of "what is now called a pragmatic dimension of language" (*ibid*.) came to be identified as increasingly important in the construction of meaning.

Therefore, in order to provide a more detailed description of the evolution of pragmatics and the theory of speech acts in the twentieth century, we shall, firstly, present some definitions of pragmatics in section 2.2 below.

2.2- Some definitions of pragmatics

Researchers in the field of pragmatics have, sometimes, demonstrated some difficulties in presenting a working definition of pragmatics, partly due to the interdisciplinary characteristic of the field and to the difficulties they face in defining and limiting the scope of pragmatics (*cf*. Levinson's 1983 attempts at defining it). However, among the various tentative definitions encountered in the literature, some are of particular interest for the purposes of the present study, since they do not constitute too broad definitions nor

excessively restrictive ones. The definitions that shall be presented in the next paragraphs can be considered as well-accepted definitions ones in the field of linguistics.

According to the *Cambridge encyclopaedia of language* (CRYSTAL 1987, p.120), pragmatics "studies the factors that govern our choice of language in social interaction and the effects of our choice on others." Blum-Kulka (1997, p. 38) defines pragmatics "in the broadest sense" as "the study of linguistic communication in context." In addition to that, the author mentions that formal definitions of pragmatics stress that "pragmatics is the science of language seen in relation to its users" (MEY 1993, p.5 *apud* BLUM-KULKA, 1997, *ibid.*); in other words, "the focus of pragmatics is on both the processes and the products of communication, including its cultural embeddedness and social consequences" (BLUM-KULKA, *ibid*.).

Levinson (1983) makes various attempts at defining pragmatics and presents the following proposal for a definition of the term.

> The most promising [definitions of pragmatics] are the definitions that equate pragmatics with 'meaning minus semantics', or with a theory of language understanding that takes context into account in order to complement the contribution that semantics makes to meaning. (LEVINSON, 1983, p. 32.)

Verschueren (1999) notes that "bolder approaches speak of pragmatics as 'the science that reconstructs language as a communicative, intersubjective and social phenomenon' " (PARRET et al 1980, p.3 *apud* VERSCHUEREN 1999, p. 262). In this perspective, pragmatics is considered by Verschueren (*ibid*.) as a field which requires a "necessary interdisciplinarity" (*ibid.*). Schmidt (1974, *apud* VERSCHUEREN *ibid*) stresses the interdisciplinary nature of pragmatics, since it has to "rely on close cooperation with other disciplines such as sociology, psychology, philosophy, logic and [others]".

Finally, Crystal (1987) notes that pragmatics is not at present a particularly coherent field of study since it focuses on the heterogeneity of discourse and on a large number of factors which affect speakers' choice of language. According to the author, it is not clear what all the factors are, "how they are inter-related, and how best to distinguish them from other recognized areas of linguistic enquiry." There would be, therefore, "several areas of linguistic enquiry. There are several main areas of overlap." (CRYSTAL, *ibid*., p. 120).

The areas of overlap mentioned by Crystal (*ibid.*) are semantics, stylistics, psycholinguistics and discourse analysis. The author believes that this overlap produces several conflicting definitions of the scope of pragmatics. In addition to that, Crystal mentions

that textbooks on pragmatics present a diversity of subject matter and some conflicting orientations and methodologies, which, however, represent a diversity of opinion which can be "a sign of healthy growth in a subject." This demonstrates, according to him, that "few other areas of language study have such a promising future" (*ibid.*, p. 121).

These views of pragmatics, though restricted in number, have been selected in the present work due to their comprehensiveness and effectiveness in representing what we consider the main characteristics of pragmatics. It is hoped that they will help clarify the scope of pragmatics with a view to a better understanding of the fundaments and workings of one of its by-products, *i.e.*, speech act theory, whose development and theoretical foundations shall be addressed in the next section.

2.3- Pragmatics and speech act theory: a historical and critical perspective

The first performative conceptions of language, as already mentioned briefly in Chapter 2.1, can be traced back in history as dating from the Ancient World and the Middle Ages. However, the pragmatic dimension of language has been receiving increased attention from philosophers and researchers from different fields, especially since the twentieth century. As a result, the different theories and studies that have stemmed from the work of these philosophers and theoreticians have taken different forms and focussed on different aspects of language and meaning.

Kerbrat-Orecchioni (2005) cites the work of various researchers who approached the study of language from a pragmatic perspective. For example, the author cites Bally (1932), who proposed a separation of the contents of every utterance into a "*modus*" and a "*dictum*", what we can presently see as a parallel to Searle's (1969) distinctions between illocutionary value and propositional content (*cf.* KERBRAT-ORECCHIONI, *ibid.*, p. 18); and Gardiner (1989), whose work in the 1930s set the fundaments of language as a means of influencing others, of the cooperative character of discourse, and highlighted the importance of bringing speech acts back to their original contexts, among others (*cf.* KERBRAT-ORECCHIONI, *ibid.*).

Other precursors of speech act theory, according to Kerbrat-Orecchioni (*ibid.*), include Benveniste (1966), who identified three modalities, namely, assertive, interrogative and imperative propositions, which, according to the latter, "reflect the fundamental behaviour of the man who speaks and acts through his discourse" (BENVENISTE, 1966 *apud* KERBRAT-ORECCHIONI *ibid.* , p. 17). These modalities were later re-named by Jakobson (1962) as referential, expressive and conative (*cf.* KERBRAT-ORECCHIONI, *ibid.*).

Another precursor of speech act theory is Morris (1938), who, according to Verschueren (1999), attempted to "outline a unified and consistent theory of signs or semiotics, which would embrace everything of interest to be said about signs." This theory would encompass the contributions of various fields, such as linguistics, philosophy, anthropology, biology, and others. Morris thus distinguished three distinct fields of study, described by Blum-Kulka (1997) as follows.

> *Syntax*, the study of the 'formal relations of signs to one another'; *semantics*, the study of 'the relations of signs to the objects to which the signs are applicable' (their referents); and *pragmatics*, the study of 'the relation of signs to interpreters.' (MORRIS, 1938, p. 6 *apud* BLUM-KULKA *ibid*., p.38.)

Therefore, as noted by Verschueren (*ibid.*), Morris introduced the "classical" definition of pragmatics, as a component of "the emergence of semiotics as a philosophical reflection on the 'meaning' of symbols […]" (*ibid.*). Philosophy has provided, according to the author, "some of the most fertile ideas in pragmatics" (*ibid.*, p. 256) and the main contributors to the development of the pillars of pragmatics were the philosophers Wittgenstein, Austin, Searle, and Grice.

In the early twentieth century Wittgenstein developed his atomistic doctrine of meaning: sentences were pictures or models of things in the world. This was taken up by the logical positivist school, who declared sentences meaningless if they didn't state some verifiable fact. Wittgenstein, at a later stage, renounced his former position and, in *Philosophical Investigations* (published post-morten, in 1953) , he advocated and elaborated the view that the meaning of language lies in the realm of language *use* and that utterances could be used *to do* many different things. The "different things that could be done with language" would correspond to one of the fundamental concepts introduced by Wittgenstein, namely, *language games*.

In the following passage, the philosopher explores the concept of "language games":

> How many types of sentences are there? Statements, questions and commands, maybe? – There are innumerable ones of such types: innumerable different types of uses of what we call "signs", "words", "sentences". This plurality is not fixed, not forever; but new types of language, new language games, as we could say, are born and others grow old and are forgotten. […]
>
> The term "language game" should highlight the fact that speaking of language is one part of an activity or of a form of life. […] (WITTGENSTEIN, 1953/2000, §23; translation mine.)

Language games, therefore, are directly associated to the notion that "meaning is use" *i.e.*, words are defined through their use in *social praxis*, which is in opposition to the former conception which views the meaning of words as related to the reference to or association with the objects they designate or to the thoughts or mental representations that people may associate with them.

Thus, Wittgenstein's (1953/2000) *non-representationalist* (*cf*. MARTINS, 2000, p.21) approach to language represented a turning-point to the conceptions of meaning which prevailed at a time when semiotics had just started to be the focus of attention. According to Martins (*ibid*.), in the representationalist view of language and meaning, "words are of secondary importance in the search for essence since they themselves contain the essence of what they designate and, also, because words owe their existence and behaviour to other things, *i.e.*, their meanings" (MARTINS, *ibid*.; translation mine).

Wittgenstein (*ibid*.), however, introduced the idea that words are not primarily representative (*cf*. MARTINS, *ibid.*, p. 27) and that language, rather than an instrument of representation, is an inseparable part of the innumerable activities which we perform in our daily lives. In addition to that, since the bond between language and human activity is inseparable, Wittgenstein argued that the meaning of phrases could not be understood as a property of phrases as independent of their uses in specific contexts, thus reinforcing his non-representationalist approach to language and meaning.

The discussions of language and meaning found in the *Philosophical investigations*, however, represent, according to Martins (*ibid*.), a step further from the slogan "meaning is use" since they lead us to reflect upon more general issues in language studies, such as language learning, the production and understanding of phrases, and the view of language as a rule-governed phenomenon [14] (*cf*. MARTINS, *ibid*., p. 29; translation mine).

---

[14] Language as a rule-governed phenomenon can be explained by the notion that "in any rule-governed activity, what determines the meaning of the ruled is the fact that the latter are put into practice" (MARTINS 2000, p. 38). In the case of Wittgenstein's language games, rules are put into practice through "customs which people do not

All of these ideas, which have stemmed from Wittgenstein's (*ibid*.) observations about the language games that people engage in and the notion that people do different things with language ("we do the most diverse things with our sentences"; *cf*. WITTGENSTEIN, *ibid.*, §27) have opened a new territory for explorations in language studies. However, other philosophers have also contributed to widening the scope of the investigations into language from a pragmatic perspective and to providing a framework for studying, systematically, different aspects involved in language use.

Blum-Kulka (1997, p. 42) notes that the work of philosophers of language like John L. Austin (1962) and John R. Searle (1969; 1975) offered "the basic insight […] that linguistic expressions have the capacity to perform certain kinds of communicative acts" and that such speech acts "are the basic units of human communication." According to Blum-Kulka (ibid.) Austin "laid the foundations for what became known as standard speech act theory" since he moves to a general theory of communicative actions, namely speech acts.

Levinson (1983), summarises the basic propositions encountered in Austin's book *How to do things with words* (1962) by arguing that Austin

> set about demolishing […] the view of language that would place truth conditions as central to language understanding. […] He noted that some ordinary language declarative sentences, contrary to logical positivist assumptions, are not apparently used with any intention of making true or false statements. […] According to Austin, […] they are not used to just *say* things, *i.e*., describe states or affairs, but rather actively to *do* things. […] Austin termed these peculiar and special sentences, and the utterances realized by them, performatives, and contrasted them to statements, assertions and utterances like them, which he called constatives. (LEVINSON, 1983, p.228-229.)

Austin, thus, introduced a novel view of language in relation to the historical context of logical positivism. However, as noted by Leech (1983, p. 176), Austin concluded that in all regular utterances, whether they have a performative verb or not, there is both a "doing" element and a "saying" element; and this led him to shift to a distinction between locutionary acts, illocutionary acts and perlocutionary acts.

A locutionary act is described by Austin (1962) as "roughly equivalent to uttering a certain sentence with a certain sense and reference"; illocutionary acts are described as "utterances which have a certain (conventional] force", and perlocutionary acts are "what we bring about or achieve *by* saying something" (AUSTIN, 1962, p. 109).

---

generally stop to reflect upon" (MARTINS *ibid*.), and which produce "no consensus of opinion but of forms of life" (*ibid*.).

In addition to that, Austin introduced the notion of "uptake", which originates in the fact that illocutionary acts, as well as perlocutionary acts, have direct and in-built consequences. Therefore, according to Austin (*ibid*., p. 116),

> […] The illocutionary act as distinct from the perlocutionary act is connected with the production of effects in certain senses:
>
> (I) Unless a certain effect is achieved, the illocutionary act will not have been happily, successfully performed. This is not to say that the illocutionary act is the achieving of a certain effect. […] An effect must be achieved on the audience if the illocutionary act is to be carried out. […] generally the effect amounts to bringing about the understanding of the meaning and of the force of the locution. So the performance of an illocutionary act involves the securing of *uptake.*

Austin dedicated the last part of his book to an attempt at classifying verbs. He argued that he needed "a list of illocutionary forces of an utterance" in "more general families of related and overlapping speech-acts" (AUSTIN *ibid*., p. 150). He thus distinguished five "very general classes" of utterances, which he "classified according to their illocutionary force", namely verdictives, exercitives, commissives, behavitives and expositives[15], but noted that he was "far from equally happy about all of them" (AUSTIN *ibid*., p. 151).

Leech (1983) notes that Austin appeared to assume throughout that verbs in the English language correspond one-to-one with categories of speech act. In this respect, Leech (*ibid*.) also notes that Searle (1979 [1975]), who proposed a similar classification, "expressively dissociated himself from Austin's assumption of such correspondence between verbs and speech acts […]. Nevertheless, […] Searle is thinking in terms of illocutionary verbs" (LEECH 1983, p. 176-177). Searle's classification of illocutionary acts consisted of five types: representatives, directives, commissives, expressives and declarations[16]. This

[15] Austin (1962, p.151-152) describes the classes of utterance he proposed as follows. *Verdictives* "are typified by the giving of a verdict [...] by a jury, arbitrator, or umpire; they may be, for example, na estimate, reckoning, or appraisal". [...]. *Exercitives* "are the exercising of powers, rights, or influence. Examples are appointing, voting, ordering, urging, advising, warning, etc." *Commissives* are "typified by promising or otherwise undertaking; they *commit* you to doing something, but include also declarations or announcements of intention […]." *Behabitives* are "a very miscellaneous group, and have to do with attitudes and social behaviour. Examples are apologizing, congratulating, commending, condoling, cursing and challenging." *Expositives* are "difficult to define. They make plain how our utterances fit into the course of an argument or conversation, how we are using words, or, in general, are expository. Examples are 'I reply', 'I argue', 'I concede', […] we should be clear from the start that there are still wide possibilities of marginal or awkward cases, or overlaps."

[16] A summarised description of Searle's five main types of illocutionary act is provided by Blum-Kulka (1997). The author describes them as follows. *Representatives* are utterances that describe some state of affairs ("The sun rises in the east') by asserting, concluding, claiming, etc. *Representatives* commit the speaker to the truth of the proposition expressed. *Directives* are utterances used to get the hearer to do something, by acts like ordering, commanding, begging, requesting and asking (questions constituting a sub-class of directives. […] *Commissives* are utterances that commit the hearer to doing something, and include acts like promising, vowing, and pledging alliance. *Expressives* include acts used to express the psychological state of the hearer, such as thanking, apologizing, congratulating and condoling […]. *Declarations* are utterances which effect a change in some, often

classification has also been criticised by other researchers, as noted by Blum-Kulka (1997, p. 43), who have argued against the "principles of classification" (*cf*. BACH and HARNISH, 1979, *apud* BLUM-KULKA, *ibid.*), Some researchers have criticised "Searle's claim that speech acts operate by universal pragmatic principles" (ROSALDO 1990 and WIERZBICKA, 1985 *apud* BLUM-KULKA, *ibid.*). The authors demonstrate the extent to which speech acts vary across cultures and languages in their conceptualization and modes of verbalization.

Leech (*ibid.*), for example, claims that Searle's classification of speech acts allows that illocutionary force may be expressed by a number of "illocutionary-force indicating devices" (IFIDS) including intonation, punctuation, and others, as well as performative verbs, but in practice, the use of devices other than performatives is not developed or illustrated in his work.

In this respect, we could add the observations put forward by Stubbs (1983), who notes that

> […] it would be quite wrong to think that just adding an IFID to an utterance makes the illocutionary force explicit without otherwise changing its meaning. For example, the following two sentences are not normally used to mean the same thing. They are certainly not interchangeable in the same social or discourse contexts, and would have different responses:
>
> 8.38 I'll come here tomorrow.
> 8.39 I hereby promise that I'll come here tomorrow.
> […] Speakers do not use extra words without reason: there are no true paraphrases without stylistic changes. (STUBBS, *ibid*., p. 157.)

Other distinctive features of Searle's classification are the importance of specific contextual conditions and their characteristic as constitutive of the different illocutionary forces performable. As Blum-Kulka (1997, p.44) notes, "the grouping of speech acts is closely tied to the set of preconditions proposed for the performance of speech acts". The conditional parameters that Searle proposed are "propositional content" (*i.e.*, "specifying features of the semantic content of the utterance […]"), "preparatory conditions" (*i.e.*, "specifying the necessary contextual features needed for the speech act to be performed […]"), "sincerity conditions" (*i.e.*, "specifying the speaker's wants and beliefs […]") and "essential condition" (i.e., the convention by which the utterance is to count as an attempt to

---

institutionalized, state of affairs. Paradigm examples are christening a baby, declaring peace […] (the types of acts included originally in Austin's 'performatives'). (*Cf.* BLUM-KULKA, *ibid*., p. 43.)

get the hearer to do something [for requests[or as an undertaking to remedy a social imbalance [for apologies]) (*cf*. BLUM-KULKA, *ibid*. p. 44). Searle also elaborated the theory of indirect speech acts (1975) through which he suggested

> a basis for a solution to the problem of systematicity: specific conventions linking indirect utterances of a given speech act type with the specific preconditions needed for the performance of the same act. This link is clearest in the case of directives: conventions of usage allow one to issue an indirect request by questioning the preparatory condition of the hearer's ability to carry out the act (*'Can you do it?'),* or by asserting that the sincerity condition obtains (*'I want you to do it.'*). (BLUM-KULKA, *ibid*., p.45.)

As Blum-Kulka (*ibid*.) notes, although Searle (1975) claims that the phenomenon is not specific to directives, examples for other speech acts are much more difficult to find. Furthermore, Searle (*ibid.*) argues that the interpretation of indirect speech acts is governed by the Gricean principle of cooperation and by conversational maxims, as well as by speech act conventions of use. However, other researchers, like Sperber and Wilson (1986) disagree. Sperber and Wilson (*ibid.*) claim that the principle of relevance supersedes all others (SPERBER and WILSON *apud* BLUM-KULKA, *ibid*., p.46) and Blum-Kulka (*ibid*.) argues that "the issue of interpretation – of how interactants match information encoded by the utterance with relevant features of the co-text and context – is far from being resolved."

Leech (1983) provides further critical comments about the issues encountered in Searle's categories of speech acts. Leech demonstrates that *a priori* categories, in fact, represent a problem to the description of natural language in use. He argues that the categorical distinctions we make while using our vocabulary do not exist in reality:

> […] we should no more assume that there are in pragmatic reality distinct categories such as orders and requests than that there are in geographical reality distinct categories such as puddles, ponds, and lakes.
>
> […] [In fact] (a) no one has no right *in advance* to assume that such categories exist in reality (although one might discover them by observation); and (b) that in actuality, when one *does* observe them, illocutions are in many respects more like puddles and ponds than like monkeys and giraffes: they are, that is to say, distinguished by continuous rather than by discrete characteristics. (LEECH, 1983, p. 177.)

Leech (*ibid*., p. 178) thus emphasises his criticism of Searle's categories of speech acts and one of the (then) current beliefs that "since language is from the pragmatic point of view a societal phenomenon, there is […] reason to suppose that it reflects precisely the distinctions that do exist in social reality. In Leech's (*ibid*.) opinion, we "cannot assume in advance such an isomorphism between language organization and social organization […]". The author

further maintains his opposition to Searle's ideas by claiming that "the only way of justifying such a view would be to make independent studies of the use of language, and of the way English (or some other language) *describes* the use of language […]" (*ibid.*). However, even in such conditions, "the evidence so far presented suggests that in general they are not [homologous]" (*ibid.*, p. 179).

Therefore, it could be argued that Leech's observations are noteworthy since they reflect the need for linguistic observations which may offer new insights into the language, rather than the adoption of *a priori* classifications. This is a fundamental distinction that permeates the present research since the classification of the utterances found in the corpus of YYRs is derived from the data encountered in actual instances of language in use, rather than categories defined in advance. The present study thus consists of data-driven[17] analyses of utterances in real-life discourse.

In addition to the purely theoretical discussions presented here, special attention has been paid to the typology and nomenclature involved in the scope of the present study, since these may vary according to the different authors (especially in the case of the terminology and typology pertaining to the field of phraseology, since this is a novel area of study). However, the theories, terms and authors discussed in the present research can be considered the most representative ones in this field in the current literature.

2.4- Speech act theory and natural spoken discourse

2.4.1- Affirmative responses in a pragmatic perspective: the politeness principle and the preference for agreement in conversation

One of the areas of study that have stemmed from speech-act theory and which appears to offer an important contribution to the present study is the **politeness principle**, postulated by Brown and Levinson (1978, *apud* GOODY, 1978 [1987]) and Leech (1983). The politeness principle, according to Brown and Levinson (1978, p. 5), "is of quite different

[17] A data-driven analysis of language can be described as an analysis of the data contained in a corpus of text (spoken or written) which have not been previously annotated, *i.e.*, classified into linguistic categories (*cf.* chapter 3.1.1 for further discussions).

status" from Grice's (1975) cooperative principle[18], since they consider Grice's maxims as background presumptions, instead of "statements of regular patterns in behaviour" (BROWN and LEVINSON, *ibid*.). As such, they account for the inferencing process in which tokens[19] of apparent uncooperative behaviour tend to be interpreted as in fact co-operative at a "deeper level", in what they name as a "*reductio*" (*ibid*.).

In Brown and Levinson's opinion, the cooperative principle defines an "unmarked or socially neutral (indeed asocial) presumptive framework for communication; the essential assumption is 'no deviation from rational efficiency without a reason." (BROWN and LEVINSON, *ibid*.)

Leech's (1983) work on politeness, however, differs from Brown and Levinson's since he argues for a politeness principle with six maxims in "interpersonal rhetoric", in addition to Grice's maxims of the cooperative principle. Leech's scheme thus extends Grice's framework, presenting the maxims of tact, generosity, approbation, modesty, agreement and sympathy (*cf*. LEECH 1983, Chapters 4, 5 and 6). Brown and Levinson (1978 [1987]) criticise Leech's work in the 1987 reissue of their work, presenting various reasons for that, among which are the following:

> [firstly], if we are permitted to invent a maxim for every regularity in language use, not only will we have an infinite number of maxims, but pragmatic theory will be too unconstrained to permit the recognition of any counter-examples; [secondly], the distribution of politeness [...] is socially controlled [...]; [thirdly], every discernible pattern of language use does not, *eo ipso*, require a maxim or principle to produce it. [...] (BROWN and LEVINSON 1978/1987, p. 4-5).

In addition to that, Brown and Levinson mention various sources which have criticised the Gricean account of communication on grounds of its cultural bias. These are related to ethnographic and sociolinguistic findings which indicate that "cultural notions of personhood are sufficiently different to make the Western emphasis on the intentional agent

---

[18] Grice's cooperative principle is summarised by Yule (1996) as follows. 'The concept of there being an expected amount of information provided in conversation is just one aspect of the more general idea that people involved in a conversation will cooperate with each other. […] In most circumstances, the assumption of cooperation is so pervasive that it can be stated as a cooperative principle of conversation and elaborated in four […] maxims' (YULE 1996, p. 37).

Grice defines the cooperative principle in the following manner. "Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged". The maxims are 'quantity', 'quality', 'relation', and 'manner' (GRICE, 1975 *apud* YULE *ibid*.)

[19] Richards, Platt and Platt (1992, p. 390) define "token" as a term used in a distinction that is "sometimes made between classes of linguistic items (*e.g*. phonemes, words, utterances) and actual occurrences in speech or writing of examples of such classes. The class of linguistic units is called a **type** and examples or individual members of the class are called **tokens**. For example, *hello, hi, good morning* are three different tokens of the type 'Greeting'."

of very dubious application" (BROWN and LEVINSON *ibid*., p. 9) in Pacific island societies, for example, as expressed in their linguistic behaviour.

Brown and Levinson (*ibid*., p. 15) argue that "there can be no universal framework of communication based on intention recognition". This would, therefore, call for a shift in emphasis from what is said to what is implicated. In this process the notion of face, *i.e.*, 'an individual's self-esteem' (*ibid*., p. 2), is central to the adoption of measurements that seek the maintenance of politeness strategies in conversation.

Brown and Levinson (*ibid.*) divide politeness strategies into three groups: positive politeness ("roughly the expression of solidarity"), negative politeness ("roughly the expression of restraint") and off-record politeness ("roughly the avoidance of unequivocal impositions"). Their uses are tied to "the relationship between speaker and addressee and the potential offensiveness of the message content" (*ibid*.).

In the case of sociolinguistics, Brown and Levinson add, "the theory argues for a shift in emphasis from [...] the usage of linguistic forms to an emphasis on the relation between form and complex inference [...]," while in the case of linguistic pragmatics "a great deal of the mismatch between what is said and what is implicated can be attributed to politeness [...]," (*ibid.*) *i.e*., the representational functions of language "should be supplemented with attention to the social functions of language, which seem to motivate much linguistic detail" (*ibid.*, p. 3).

The politeness principle could, perhaps, be partly associated with the occurrences of *yeah* and *yes* responses in English, due to their potential role of counteracting aggression. Brown and Levinson (*ibid*., p.1) argue that

> the problem for any social group is to control its internal aggression while retaining the potential for aggression both in internal social control and, especially, in external competitive relations with other groups [...]. In this perspective politeness, deference and tact have a sociological significance altogether beyond the level of table manners [...]; politeness [...] presupposes that potential for aggression as it seeks to disarm it, and makes possible communication between potentially aggressive parties.

Yeah and *yes* responses (henceforth referred to as YYRs), as any other type of affirmative response, would thus be valid tools in the performance of politeness strategies. However, it should be noted that this does not mean that this is their sole function in language, nor that politeness implies responding affirmatively. Many other politeness strategies are found to be used in speech acts of disagreement, refusal, or other negative ones, for instance.

The case of YYRs, thus, runs close to the issues raised by Brown and Levinson in the politeness principle, since the understanding of their realisation in English involves an understanding of the 'deviations' from 'rational efficiency', and the different implicatures

they may contain. They are ways of expressing various meanings, all of which are either upgraded or mitigated through phrases or clauses that frequently include or occur in conjunction with the words *yeah* or *yes*. *Yeah* and *yes* thus may serve as atenuators in face threatening acts (FTAs; in negative politeness) or as heads in cases of positive politeness or off-record strategies, if we are to consider *yeah* and *yes* responses in the perspective of the politeness principle (*cf*. the instances of *yeah* and *yes* responses in the corpus under analysis in the present research). Language users might, therefore, use such strategies more or less frequently according to context, as shall be discussed later in the present research.

2.4.2- Preference organisation and types of responses

A study of the subtleties and nuances of YYRs and the various speech acts they may perform leads us to an investigation of how they are structured in discourse.

Here, it would be of interest to note Brown and Levinson's (1978[1987]) remarks about the pertinence of studies in preference organisation, a term used by conversation analysts. It refers to

> the phenomenon that after specific kinds of conversational turn, responses are often restrictly non-equivalent: one kind of response, termed the preferred, is direct, often abbreviated and structurally simple, and typically immediate; in contrast, other kinds termed dispreferred are typically indirect, structurally elaborated, and delayed. (Pomerantz 1975, 1978, 1984a). The preferred type of response is usually more frequent also, but the term 'preference' refers to the structural disposition, to the fact that conversational organisation conspires to make it easier to use the preferred type of turn, not to participant's wishes. In this sense it can be shown that there are preferences for matters as diverse as: (i) agreement (vs. disagreement); (ii) repair by self (vs. repair by other of mistake or unclarity by self); (iii) acceptances (vs. rejections) of requests and offers; (iv) answers (vs. non-answers) to questions; in addition, preferences also hold across sequence types, for example, (v) offers by A (as opposed to requests by B to A); (vi) recognition by other of self on telephone (vs. self-identification); and so on. (BROWN and LEVINSON *ibid*., p. 38).

The notion of preference is further expanded by Levinson (1983, p. 332-333), who argues that it "is not intended as a psychological claim about the speaker's or hearer's desires, but as a label for a structural phenomenon very close to the linguistic concept of markedness, especially as used in morphology."

The concept of markedness was originally developed by linguists of the Prague School and can be defined as the tendency of one member of an opposition (between two or more members) to be "felt to be more usual, more normal, less specific than the other" (COMRIE, 1976 *apud* LEVINSON, *ibid.*).

Levinson's argument based on Comrie's (1976) definition is that "in a similar way preferred (and thus unmarked) seconds (*i.e.*, the second part of an adjacency pair) to different and unrelated adjacency pair first parts have less material than dispreferreds (marked seconds), but beyond that have little in common." (LEVINSON *ibid*.). He also points out that "in addition to the structural aspect of preference organisation, we will need a rule for speech production, which can be stated roughly as follows: try to avoid the dispreferred action." (*ibid*.).

What determines which types of response are preferred or dispreferred would be mostly associated with face considerations in Brown and Levinson's point of view. They argue that

> agreement is preferred because disagreement is an FTA; [...] acceptances of offers or requests [are preferred] because the alternative refusals would imply lack of consideration; [...] In the case of the preference for an offer-acceptance sequence over a request-acceptance sequence, [...] there is less face risk [in the former] because B may refuse the request [...]. Thus face considerations seem to determine which of two alternative responses after another turn will be normally associated with the unmarked, preferred turn format. (*Ibid*., p.38-39).

Pomerantz's (1975) studies into agreement mentioned in Brown and Levinson's (*ibid*.) work are also of interest in cases of conflicting requirements derived from speakers' face considerations. When reacting to compliments, for instance, speakers resorted to intermediate strategies between the preferred ones of agreement and constraint against self-praise or criticism of alter. These, according to Brown and Levinson (*ibid*, p.39), "seem to lie firmly in the realm of face-motivated behaviour". The preference for agreement would thus be related to positive face considerations, while the constraint against self-praise could be paralleled to instances of negative politeness, such as the use of honorifics, for instance.

The importance of face considerations in preference organisation can also be confirmed by the fact that it provides speakers with the necessary tool for many "face-preserving strategies and techniques," according to Brown and Levinson (*ibid*.). Preferred turns usually follow a first turn immediately and instances of delaying them, for example, may represent the recipient's reluctance to produce the referred action; also, some entire sequences of conversational turns have been found to be preferred to others. Some examples of this are the inducing of offers over the actual request, reporting facts in order to elicit other reports or offers, for example, and others encountered by Brown and Levinson (*ibid*., p. 40).

Therefore, these preferred sequences and strategies, which may also include pre-sequences and finishings in negotiations of all sorts, can be justified in terms of face saving. The face-saving awareness by speakers is further illustrated by Brown and Levinson through

the findings of Jefferson (1980 and 1984 *apud* BROWN and LEVINSON *ibid*.) about the reporting of troubles: here face implications have been found to structure the entire conversation from the point of their introduction.

Through the concepts of social felicity conditions, mitigation and upgrading introduced by Labov and Fanshell (1977 *apud* BROWN and LEVINSON, *ibid*.), Brown and Levinson predicted a scale of politeness (or 'mitigation') of indirect requests. They further concluded that the theory of speech acts "[...] can be recast in sequential terms: so-called indirect speech acts are in fact pre-sequences designed for cooperative pre-emption or tactful evasion [...]" (*ibid*., p.42).

As a consequence, they argue for the importance of a focus on what comes before and what comes next to conversational turns that realise face-threatening acts. The latter, which is focused on in their work from the point of view of its internal structure, should thus be complemented by such a focus, in their opinion.

2.4.3- Preferred and dispreferred seconds

The analysis of the structure of both first and second turns in speech reveals that interactional feedback is systematically taken into consideration over the course of a single turn's construction (*cf*. DAVIDSON, in press, *apud* LEVINSON, 1983, p. 336).

> In this sense a single turn at talk by one speaker can itself be seen to be a joint production [...]. There is also further evidence of quite different kinds which shows that a single speaker's turn is often a joint production, in that recipient's non-verbal responses are utilized to guide the turn's construction throughout the course of its production (see Goodwin, 1979a, 1981). Here, though, preference organization, in constraining the construction of second parts of adjacency pairs, can systematically affect the design of first parts. (LEVINSON, *ibid*., p. 337).

According to Levinson, preference organisation is not limited to adjacency pairs, *i.e.*, there are other turns where "turns paired less tightly than adjacency pairs, where a first part does not seem to require but rather makes apt some response or second-action chains in Pomerantz's (1978) terminology". He gives the example of a second assessment (agreement or disagreement) which is due after a first assessment, and mentions that there is a clear preference for agreement over disagreement (LEVINSON *ibid*., p. 337-338.).

However, when assessments involve complex issues of self denigration or compliment, for instance, the conversation expectations work in opposing directions and compromise solutions are usually employed by Ss: with the former the general norm is the

avoidance of criticism, while with the latter it is the avoidance of self-praise, according to Pomerantz (1978 *apud* LEVINSON 1983).

2.4.4- Preferred sequences and sequence types

An area where preference organisation routinely operates within and across turns is the organisation of repair, as demonstrated by Levinson (1983, p. 340) "the tendency for an utterance to attend to those immediately prior to it provides, for both analysts and participants, a 'proof procedure' for checking how those turns are understood."

Repair is divided into self initiated (repair by a speaker without prompting), other initiated (repair after prompting), self repair (repair done by the S of the problem or repairable item) and other repair (done by another party). The data found in Schegloff, Jefferson and Sacks (1977 *apud* LEVINSON 1983) illustrate that an important component of the repair apparatus is a set of preferences setting up a rank ordering. The preference ranking is, from the most frequently used to the least used resource: 1- self initiated self repair in one's own turn; 2- self initiated self repair in transition space; 3- other initiation of self repair in the next turn; 4- other initiated other repair in the next turn.

This led Levinson to argue that the system is actually set so that there will be a tendency for self-initiated self-repair and that in the rare event of other-repair occurring, it is followed by 'modulators' such as 'I think', or prefaced by elements such as 'y' mean'. Therefore, "the repair apparatus as a whole is strongly biased by a preference for self-initiation of repair and by a preference for self-repair over repair by others. As a consequence preference organization governs the unfolding of sequences concerned with repair." (LEVINSON, 1983, p. 342).

As far as sequence types are concerned, Levinson demonstrates that preference can play an important role in determining which speech acts are chosen by speakers. He mentions the case of pre-requests, which can be operated by the prompting of an offer and which have been found to be preferable to performing a request.

The example he provides of a pre-request, which is reproduced here, is also a good example of the role of YYRs in such preferred sequences. Their role in the negotiation of meaning is demonstrated here.

(80)
C: Hullo I was just ringing up to ask if you were going to Bertrand's party
R: *Yes* I thought you might be
C: Heh heh
R: *Yes* would you like a lift?
C: Oh I'd love one
(LEVINSON, 1983, p. 343).

The same is true of telephone calls, as mentioned earlier, where self-identification is usually dispreferred, whereas the preference is for callers to provide the minimal cues they judge sufficient for recipients to recognise them.

In Levinson's (1983, p. 345) words,

> preference organization thus extends not only across alternative seconds to first parts of adjacency pairs, but backwards into the construction of first parts, forwards into the organization of subsequent turns, and also across entire alternative sequences, ranking sets of sequence types.

2.5- Conversational sequences, *yeah* and *yes* responses and indirect speech acts

2.5.1- Pre-sequences - definition; turn location, turn position, and the characterisation of sequences

Pre-sequences, according to Levinson (1983), can refer either to a certain kind of turn (referred to as pre-s) or to a certain kind of sequence (referred to as pre-sequence) containing that type of turn. A pre-s is often "built to prefigure the specific kind of action that [it] potentially precede[s]. [...] It is a turn that occupies a specific slot in a specific kind of sequence with distinctive properties." (LEVINSON *ibid.*, p. 346). Some examples are: pre-announcements, pre-requests, pre-arrangements for future contact, pre-invitations, pre-offers, and the like.

The structure of such sequences would be as follows:

> a) T1 (position 1): a question checking whether some precondition obtains for the action to be performed in T3;
> T2 (position 2): an answer indicating that the precondition obtains, often with a question or request to proceed to T3;
> T3 (position 3): the pre-figured action, conditional on the 'go ahead' in 'T2;
> T4 (position 4): response to the action in T3;
> b) *distribution rule*: one party, A, addresses T1 and T3 to another party, B, and B addresses T2 and T4 to A.
> (LEVINSON, *ibid.*, p. 346-347).

Schegloff's (undated, *apud* LEVINSON 1983, p. 348) distinction between turn location and turn position is also important here. Turn location is "the sheer sequential locus of a turn in a sequence by a count after some initial turn", whereas position refers to "the response to some prior but not necessarily adjacent turn. Thus a second part of an adjacency pair separated from its first part by a two-turn insertion sequence will be a fourth turn but second position."

Levinson exemplifies the fact that the characterisation of each position is possible (independently of absolute location in a sequence of turns) through the case of pre-announcements. He claims that

> the design of the turn in position 1 is crucial: for it is on the basis of this that the recipient must decide whether or not he already knows the content of the announcement, and thus should abort the sequence. Hence the prefiguring of the syntactic frame of the announcement [...] is a very useful clue to the recipient (LEVINSON *ibid*., p.352.)

In the case of YYRs, we can note here the importance of their role (as would, also, be the case of any other affirmative or negative responses) in the negotiation of meaning in conversational sequences, a preoccupation which is at the basis of the present study. In the following example, provided by Levinson (1983, p.348) to illustrate the importance of the pre-sequence sub-type pre-request, we can notice the use of *yes*-responses in T3 (response to repair in T2), in T6 (position 2 response to the pre-request in position 1), and in T7 (in acknowledgement of T6), followed by the actual position 3 request (which might have been followed by a position 4 YYR, in a possible response not shown in the data).

(95)

T1	C: ... Do you have in stock please any L.T. one eight eight? ((POSITION 1))
T2	R: One eight eight ((HEARING CHECK))
T3	C: Yeah ((CHECK OKAYED))
T4	R: Can you hold on please ((HOLD))
T5	C: Thank you ((ACCEPT))
T6	R: *Yes* I have got the one ((POSITION 2))
T7	C: *Yes*. Could I _ you hold that for H.H.Q.G. please ((POSITION 3)).

Therefore, the use of a *yes* response appears in the example above in a close relationship with matters of preference organisation and preferred sequences in conversation, as discussed in the previous sections.

2.5.2- Indirect speech acts from a conversational sequence perspective

Levinson's (1983) critical analysis of indirect speech acts can be considered of great usefulness to the present study since it provides a better understanding of the motivations behind speakers' choices in the sequential turn-taking system in conversation, where YYRs may and usually do play an important role. He analyses requests and pre-requests with a view to re-analysing indirect speech acts, and justifies his choice by the fact that requests are "the variety of indirect speech act which has received the most attention" (*ibid*., p.357).

He noted that the use of pre-requests is motivated by "the preference ranking which organises responses to requests themselves." They allow speaker 1 (S1) to verify whether a request is likely to succeed and, "in cases of doubt, pre-requests are to be preferred to requests."(*ibid*.). Also, pre-requests usually contain a precondition that is associated with the refusal of that request. For example, in conversation they question the recipient's abilities, while in service encounters they check whether goods are in stock or not, once these are the preferred grounds for refusals in such cases. "What is checked in the pre-request is what is most likely to be the grounds for refusal [...]" (*ibid*., p.358).

Therefore, as position 1 turns they check ("and are therefore generally questions", [*ibid*.]) the most likely grounds for rejection as well as avoid an action that would obtain a dispreferred second. In addition to that, they may serve to avoid overt actions, such as a request in itself or an offer; *i.e*., pre-requests may be an effective clue to the obtaining of the desired action by S2, once he/she would prefer "to provide it without more ado" (Levinson 1983, p.360), rather than offer to do it or prompt the request by S1. To that, Levinson encounters some parallels in the preference for recognition in telephone calls, the preference for self-initiated self-repair or for embedded (or covert or implicit) correction over exposed correction in repair, as mentioned earlier.

He also believes that there may be a preference for the avoidance of requests altogether, and that the following preference ranking operates over three kinds of sequences (excluding those that are aborted when preconditions are not met):

(i) most preferred:	Position 1: (pre-request)
	Position 4: (response to non-overt request)
(ii) next preferred:	Position 1: (pre-request)
	Position 2: (offer)
	Position 3: (acceptance of offer)
(iii) least preferred:	Position 1: (pre-request)
	Position 2: (go ahead)
	Position 3: (request)
	Position 4: (compliance)

(LEVINSON, *ibid*., p. 361).

In a parallel to the case of pre-announcements, where position 1 turns are often carefully formulated to provide information about the upcoming announcement and sometimes formulated to obtain guesses in position 2 (what would otherwise appear in position 3), pre-requests can be built to invite position 4 responses (through the provision of enough information, the use of markers of interaction, pessimism, pre-verbal *please*, and others (*cf*. BROWN and LEVINSON 1978[1987], p.320).

Indirect speech acts, Levinson concludes,

> are position 1 turns (...) formulated so as to expect position 4 responses in second turn. (...)They can be formulated so as to project certain conversational trajectories (...). In preference organisation a systematic preference for the avoidance of some sequences altogether (...) provides a motivation for the collapse of the four-position sequence into the two-position sequence consisting of a position 1 turn followed by a position 4 turn (...); we can expect position 1 turns to be expressly formulated to get position 4 turns in second turn - and hence for pre-requests of this sort to contain special markers (including *would, could, not, please*, etc.). (LEVINSON, 1983, p.363, 364).

Levinson's purpose in these analyses is to demonstrate the way in which conversational analysis insights can be useful to the solution of linguistic problems and to the study of linguistic form. The cases of requests, pre-requests and other pre-sequences investigated by Levinson are only a few of the many conversational features observable in the language. As Levinson observed, his choice for requests was due to the fact that these have received much attention in the literature; however, in the present study an option has been made for investigations in an area which has not received much attention in the literature, as already noted in the Introduction (2.1), *i.e*., the case of *yeah* and *yes* responses. These are frequently found in sequences of the types analysed by Levinson and which were described here in this chapter; however, Levinson's main focus was on the initiation moves (IMs) that served as pre-sequences or as the head act in the sequences shown. In the present study, however, focus has been displaced to the reverse side of the coin: how S2 interacts with S1 whenever a YYR is used in conversation, in whatever position they appear. Another difference here is that the present study is a corpus-driven analysis, *i.e*. an analysis of the data contained in a corpus of (spoken) text which have not been previously annotated (or classified) into linguistic categories (*cf*. Chapter 3.1.1 and Chapters 5 and 6 for further discussions). In other words, the present research is different in nature from conversation analysis.

Levinson's insights into conversation analysis have been reproduced in this Chapter with the purpose of illustrating the fact YYRs operate, among other things, within the most

preferred sequences in various speech acts, such as requests and others. The present study, however, is not a conversational analysis of their occurrences in English, but, rather, a corpus analysis.

3- FINDING PATTERNS IN LANGUAGE: CORPUS LINGUISTICS AND STUDIES INTO SPOKEN ENGLISH

> "The discourse of applied linguistics would have little grounding in reality if it could not be measured against change in method and practice […]. It so happens that corpus linguistics both represents cutting-edge change in terms of scientific techniques and methods, and presents us with dilemmas that arise from the humanistic contexts in which an (apparently) detached technology operates. Corpus linguistics probably also foreshadows even more profound technological shifts that will impinge upon our long-held notions of education, the roles of teachers, the cultural context of the delivery of educational services and the mediation of theory and technique as the twentieth century becomes history."
> Michael McCarthy - *Issues in applied linguistics* – 2001, p. 125.

3.1- Sense and structure as patterns of language in corpus-driven studies

3.1.1- Corpus linguistics: focus on naturally-occurring language

The new possibilities that have arisen due to the use of the computer have enabled researchers to develop new approaches to language studies. Notably, corpus linguistics, among other applications, can be considered one of the most comprehensive and promising technology-assisted fields. Hunston and Francis (2000, p. 14) define corpus linguistics as "a way of investigating language by observing large amounts of naturally-occurring, electronically-stored discourse, using software which selects, sorts, matches, counts and

calculates". Leech (1992 *apud* GRANGER 1998, p. 3) considers corpus linguistics not just a new computer-based methodology but "a new research enterprise". Granger (*ibid*.) argues that corpus linguistics, with its focus on performance (rather than competence), description (rather than universals) and quantitative as well as qualitative analysis, can be seen as "contrasting sharply with the Chomskyan approach"; however, the author notes that these two approaches "are not mutually exclusive". Fillmore (1992, p. 35 *apud* GRANGER *ibid*.) argues that "the two kinds of linguists need each other. Or better, […] the two kinds of linguists, wherever possible, should exist in the same body."

Furthermore, Hunston and Francis (2000, p. 14), note that one of the pioneers in this field, namely, Professor John Sinclair (1987a,b; 1991), "prioritises a method, or group of methods, and a kind of data rather than a theory" (*ibid*.), which confers a unique quality to Sinclair's work. In other words, the type of research developed by Sinclair in the field of corpus linguistics has allowed for theory to be derived from naturally occurring data which is compiled by the computer, as opposed to traditional language studies, which are characterised by the *a priori* application of theory to idealised language.

According to Hunston and Francis (*ibid*.), the data used in corpus analysis can be distinguished from the data used in other approaches in five respects, as follows.

> The data is authentic;
> The data is not selected on linguistic grounds;
> There is a lot of data;
> The data is systematically organised;
> The data is not annotated in terms of existing theories.
> (HUNSTON AND FRANCIS 2000, p. 14-15)

As Hunston and Francis point out, […] "each of these features may be stated as a principle […]" (*ibid*.) since they are the fundaments of Sinclair's approach to the use of computerised corpora.

The importance of the first feature, *the data is authentic*, can be contrasted with what grammarians have done for many years, *i.e.*, they have relied on introspection and intuition. In this respect, Sinclair (1991, p. 4) argues that

> […] the contrast exposed between the impressions of language detail noted by people, and the evidence compiled objectively from texts is huge and systematic. It leads one to suppose that human intuition about language is highly specific, and not at all a good guide to what actually happens when the same people actually use the language.

In other words, intuitions about the frequency of words in actual use are counteracted by proof obtained from the compilation of corpora by the computer and their subsequent analyses, a topic that is constantly addressed by Sinclair and other linguists (*cf*. HUNSTON,

2002; MCARTHY and CARTER, 1995; WILLIS 2003; WILLIS 1990). These researchers make it clear, however, that, although the most frequent words are more useful to learners for the acquisition of their phraseologies. Less frequent words and their phraseologies are also very important for the acquisition of the language that is used, for example, in specialist contexts, which cannot be disregarded in teaching, and in the elaboration of specific language teaching materials. These issues shall be further discussed in this chapter.

The second feature, the fact that the data at the basis of the work implemented by Sinclair is *not selected on linguistic grounds*, distinguishes his work from the work of other linguists who have also derived their data from actually-occurring discourse, such as those involved with functional grammar, but have traditionally selected actual instances of language in order to illustrate specific items. Sinclair (1991 *apud* Hunston and Francis *ibid*., p.16) notes that "this method is likely to highlight the unusual in English and perhaps miss some of the regular, humdrum patterns." Or, as Hunston and Francis (*ibid.*) describe it, "where instances of language are selected for analysis precisely because they strike the linguist as interesting, they are likely to exemplify the unusual rather than the mundane."

The third feature of Sinclair's work, *there is a lot of data*, refers to his view of quality as related to quantity in corpus studies. As Sinclair (*ibid*., 100) describes it, "[…] the ability to examine large text corpora in a systematic manner allows access to a quality of evidence that has not been available before."

In addition to that, Hunston and Francis (*ibid*., p.16) note that "the difference between looking at a lot of data and a little, is that when a lot of data is examined, conclusions as to frequency can be drawn." They exemplify this through the rephrasing of Sinclair's (1991) observations about language.

> […] firstly […] some sequences of words co-occur surprisingly often, given that every utterance or written sentence spontaneously produced is unique; secondly, and in contrast, that even so-called fixed expressions demonstrate surprising amounts of variability (*cf.* MOON 1994[20]; 1998 *apud* HUNSTON and FRANCIS, *ibid*.)

As a consequence of that Hunston and Francis (*ibid*., p. 17) argue that

> Language is not a system that is realised in actual instances, but a set of actual instances that may be regarded as construing an approximate and ever-changing system. Such a construal stems from the interpretation of hundreds of observations, made possible by the sheer amount of data available.
> One of the outcomes of using large quantities of data is that some of it may be discarded, in the sense that instances of word-play or language that is strange

[20] MOON, R. The analysis of fixed expressions in text. In: COULTHARD, M. (Ed.) *Advances in written text analysis*. London: Routledge, 1994, p. 13-27, *apud* HUNSTON, S., FRANCIS, G. *Pattern grammar: a corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins, 2000.

> because it is being used in strange circumstances, are deliberately ignored in terms of the general description of the language (Sinclair 1991:99), though they may form the focus of studies of a different kind […]. This is a different approach from that of many grammars, which concentrate on what is possible, not what is frequent.

The fourth feature, *the data is systematically organised*, relates to the fact that the data in the Bank of English has been organised with the word-form as the unit. A word-form is defined by Sinclair (*ibid.*, p. 176) as "a series of characters separated from other series by a space;" or "any unique string of characters, bounded by spaces. Hence *give, giving, gave, given* are all separate word-forms."[21] All instances of a given word-form can be searched in the corpus through the use of software that has been designed for this purpose; it also presents word-forms together with a limited amount of the preceding and following text in concordance lines. As noted by Hunston and Francis (*ibid.*, p.18), "a method that takes the word-form as the focal point in the presentation of data is bound to result in a theory of language that prioritises the word-form and its behaviour."

The last of the five features of Sinclair's approach, *the data is not annotated in terms of existing theories*, has significant implications to corpus studies of language. With annotated corpora, the software searches for a particular category, which is called "corpus-based" research by Tognini-Bonelli (1996 *apud* HUNSTON and FRANCIS, *ibid.*). Tognini-Bonelli (*ibid.*) contrast "corpus-based" research with "corpus-driven" research, which is the type of research developed by Sinclair (ibid.) and developed in the present study.

However, annotation software and annotated corpora have some disadvantages and problems, such as the fact that "[…] automatic taggers and parsers[22] have limited accuracy […]"; also, "[…] the annotation will reflect a particular theory of grammar, and the results will naturally be cast in terms of that theory" (HUNSTON and FRANCIS, *ibid.*, p.19, after Sinclair). In this respect, Sinclair argues that

> If…the objective is […] to observe and record behaviour and make generalisations based on the observations, a means of recording structure must be devised which depends as little as possible on a theory. The more superficial the better. (SINCLAIR 1987c, p.107.)

Furthermore, Hunston and Francis argue that

> […] the question of method – how to investigate the large amounts of data available in a corpus – is a crucial one to corpus linguistics, and one that no-one as yet is in a position to answer fully. What is more surprising, perhaps, is that the literature

[21] In order to define the term 'word-form', Sinclair contrasts it with the term 'lemma', which he defines as "[…] the composite set of word-forms". For example, the lemma *give* has the different word-forms "*give, gives, gave, given,* and *giving*" (SINCLAIR 1991, p.174).

[22] Taggers are defined by Hunston (2002, p. 80, 82) as "programs that assign tags" *i.e.*, "allocate a part of speech [POS] label to each word in a corpus"; parsers are defined as programs that analyse "[...] sentences in a corpus into their constituent parts, that is, doing a grammatical analysis." (*ibid.*, p. 84)

> reveals very little in the way of methodological debate. (HUNSTON and FRANCIS, *ibid*.)

Therefore, it could be argued that, in this respect, the present research consists of corpus-driven analyses that have been developed in accordance with Sinclair's approach since the data is authentic, has not been selected on linguistic grounds, contains large amounts of tokens (in fact, all the tokens for the nodes *yeah* and *yes* in the Bank of English corpus), is systematically organised and is not annotated in terms of any existing theories (*cf*. Chapter 5).

3.1.2- The importance of corpus evidence to language studies

Corpus linguistics in general and corpus-driven studies of language, have, to a large extent, been greatly influenced by Sinclair, who, according to McCarthy (2001, p. 127), has been "foremost in seeing […] profound implications of corpus linguistics for a radically different view of language as a whole." The distinctive features of Sinclair's work are summarised by McCarthy (*ibid*.). Firstly, "Sinclair's work represents the classic case of independent applied linguistics, in that he has come from practice to a new theory, not vice-versa. This […] is one of the key ways in which applied linguistics carries out its discourse and defines itself […]." Secondly, Sinclair, who engaged in the "eminently practical pursuit of writing a learners' dictionary in the early 1980s", became aware that "certain dearly held principles of language study (*e.g*. the primacy of syntax, the unpredictability and 'irregularity' of lexis) were simply no longer tenable when faced with corpus evidence". What Sinclair could verify was that lexis "was far from irregular; regular vocabulary patterns appeared everywhere in the corpus." Thirdly, idiomaticity "appeared to be ubiquitous and at least as important as syntax in the construction of meaning."[23] Therefore, McCarthy (*ibid*.) notes, far from being a minor, or marginal affair in language, "idiomatic constructions were everywhere in the corpus Sinclair was working with, especially in the patterns formed by combinations of the most frequent words in the language." McCarthy thus contrasts the high frequency of idiomatic expressions "to the quaint, infrequent idioms often associated with language teaching manuals." Finally, the identification of the high frequency of idioms "and other factors" (*cf*. McCARTHY *ibid*.) led Sinclair

> to posit a close bond between sense and structure, and to conclude that features such as collocation and particular idiomatic (in the sense of individual) but very

[23] (Sinclair 1991, p.112 *apud* McCarthy 2001, p. 127.)

> frequently occurring combinations were the real cement that held texts together. Syntax was more, as it were, an emergency repair kit for filling the occasional gaps and cracks amid the flow of idiomaticity. (McCARTHY 2001, p. 127.)

The new evidence obtained from corpora, thus, enabled Sinclair to formulate new theories and principles, which, perhaps, could be considered as the cornerstone of linguistic descriptions in the 21st century. In McCarthy's words,

> Sinclair's proposal is radical […] but it stands as a good example of how a 'neutral'[24] technology can throw up fundamental questions for theory, and how a practical, 'applied' problem, in this case writing a dictionary using computer evidence, can bounce back and challenge theory. We should not doubt that galloping technological change will bring many more such upheavals over the coming decades. (McCARTHY 2001, p. 127)

3.1.3- Phrases, sense and structure as patterns of language

The work of Sinclair, though initiated more than two decades ago, is fundamental to an understanding of the state-of-the-art in corpus studies. Some of the principles he envisaged are basic to an understanding of how corpus-driven studies can contribute to investigations into language use. Since the focus of Sinclair's work in corpus linguistics is the word, its collocates and patterns of usage, he discusses the issues of word meaning and structure to finally formulate the existence of the idiom principle, which shall also be addressed in this chapter (*cf*. 3.1.4).

The different senses of a word and the structures in which they occur in the language have, according to Sinclair, a "strong correlation". He thus proposes a definition of "structure" as including "[…] lexical structure in terms of collocations and similar patterns" whereas " 'senses of a word' includes the contribution that a word may make to a multi-word lexical item." (SINCLAIR, 1991, p. 53.) The evidence Sinclair encountered in his analyses of patterns in the corpus led him to conclude that sense and syntax seem to be strongly associated and that "that adjustment of meaning and structure is a regular feature of a language." (SINCLAIR *ibid*., p. 65.)

The correlation between sense and structure that Sinclair identified in the analyses of actual instances of the language would enable researchers to select the most typical

[24] McCarthy (2001, p. 128) notes that apparently neutral technologies are rarely neutral and that "it is in the practice of those technologies that applied linguists are often obliged to return to the basics of metalanguage and the mutual discourse with which they define their activities."

occurrences of phrases and their patterns. This would, consequently, lead to changes in lexicography, grammars and other areas. As argued by Sinclair,

> The new option opened up by the computer is to evaluate actual instances and select the most typical. A complete set of typical instances should exemplify the dominant structural patterns of the language without recourse to abstraction, or indeed to generalization. […]
>
> In the explicit theoretical statement of linguistics, grammatical and lexical patterns vary independently of each other. In most grammars, it is an assumption that is obviously taken for granted. For example, it is rare for a grammar to note that a certain structure is only appropriate for a particular sense of a word. The same goes for morphology. […]
>
> Equally, it is rare for a dictionary to note the common syntactic patterns of a word in a particular sense. (SINCLAIR, 1991, p. 103)

As a consequence of the new possibilities offered by the use of computerised corpora, present-day lexical research places emphasis on the phrases of the language, thus, bringing grammar and lexis together. Sinclair (*ibid.*, p. 104) proceeds in his findings to introduce his definition of phrases: "A phrase can be defined for the moment as a co-occurrence of words which creates a sense that is not the simple combination of the sense of each of the words."

In addition to that, Sinclair also presents a definition of structure that differentiates it from the most frequent and traditional senses found in linguistics, since it encompasses both lexis and grammar. He describes it as "any privileges of occurrence of morphemes; we do not in the first analysis have to decide whether these are lexical or syntactic – or as so often – a bit of both" (SINCLAIR, *ibid.*).

Another feature of Sinclair's research is his formulation of the hypothesis that the 'underlying unit of composition is an integrated sense-structure complex', which he uses in order to account for the view of sense and structure as units that are not independent of each other and not inseparable. Sinclair (1991, p. 104) thus introduces the issue with the following question,

> [...] Is it [...] best to hypothesize that sense and structure are inseparable? Unfortunately not. [...]
>
> If sense and structure are not independent of each other and not inseparable, then they must be associated. Here we can frame a hypothesis that can act as a substitute for the langue/parole distinction. We can postulate that the underlying unit of composition is an integrated sense-structure complex, but that the exigencies of text frequently obscure this. This position offers a sharp contrast to the atomistic model featured by most grammars [...].

Sinclair then proceeds to emphasise the new descriptive task of the linguist, *i.e.*, to identify the recurrent patterns of the language, which, in turn, leads to the identification the different citation forms. Once these are identified, the researcher can proceed to state their distinguishing features and provide explanations for "the occurrence of non-citation forms."

The author notes that a citation form "would involve a modest step in abstraction. It is also likely that many citation forms contain some syntactic variables, such as pronoun selections, which leaves a modicum of independence to the grammar." (SINCLAIR *ibid*., p.105.)

However, he points out that the association between sense and pattern (or syntax, or structure) is not one-to-one. "More than one sense can be realized by the same structure, and, in the simplest case, by the same word" (*ibid*., p. 104). Or as Hunston and Francis (2000, p. 21) describe it, "in ordinary discourse, structure does not satisfactorily distinguish sense."

As already mentioned in the preceding paragraphs, Sinclair argues that "[…] the underlying unit of composition is an integrated sense-structure complex […]" (*ibid*., p. 105), which resides in the phrases of a language, *i.e*., co-occurrences of words which are not simple combinations of the sense of each of the words" (*ibid*., p.104). This can be well illustrated by the investigations that he made into phrasal verbs. Sinclair notes that

> Each sense of the phrase is co-ordinated with a pattern of choice that helps to distinguish it from other senses. Each is particular; it has its uses and its characteristic environment. [...]
>
> If *set in train* always occurs together in this sequence when it has the obvious meaning, then the three words constitute one choice. As soon as learners have appreciated that each phrase operates as a whole, more or less as a single word, then the difficulty disappears and they have a new word *set in train*. [...] Once it is clear that what matters is the meaning of the phrase as a whole, then any recollection of the independent meanings of the constituent words will reinforce the phrase meaning. (SINCLAIR, *ibid*., p. 78,79).

Further examples of the importance of the phraseology of the language to the creation of the different senses of a word can be found in Stubbs (2002, p. 3,4) through the cases of 'round' and 'table' and their phraseologies. According to the author, the two words have possible meanings when presented in isolation (*i.e*., "circular" and "a piece of furniture with a flat top, which people can sit at, so that they can eat, write, and so on", respectively). The phrase *round table* has one meaning "which is simply due to the combination of these individual meanings: something which is both 'round' and 'table'. However, it is also used in longer phrases such as "*round table talks*" (*i.e*., "a group of people, with interests and expertise in some topic, are meeting as equals to discuss some problem"). Stubbs notes that the meaning in "*round table talks*" is culturally dependent and that those words can mean quite different things in other phrases, such as in "*a round number*", "*a table wine*", "*a timetable*". This is due to the fact that "most everyday words have different uses and different meanings. Indeed, in isolation, some words seem to have so many potential meanings that it is difficult to see how we understand running text at all," the author notes. However, Stubbs points out that words do not occur in isolation, but in longer phrases, such as "they sat *round*

the table", "they ran *round* the table", "they came *round* to my house", "they came *round* to my way of thinking", "a *round* dozen", "a *round* of applause", and other phrases from which people can normally extract the relevant meaning of the word "round". The author argues that "our knowledge of a language is not only a knowledge of individual words, but of their predictable combinations, and of the cultural knowledge which these combinations often encapsulate" (*cf*. STUBBS, *ibid*., p. 4).

Some of the issues related to the identification of meaning which have been posed by Sinclair are also the subject of Stubbs' (2002) discussions. According to Stubbs, "Sinclair puts forward the hypothesis that units of meaning are 'largely phrasal', that only a few words are selected independently of other words, and that 'the idea of a word carrying meaning on its own [can] be relegated to the margins of linguistic interest." (*ibid*., p. 63). Examples of the latter case are cited by Stubbs (*ibid*., *apud* SINCLAIR 1991, p. 82) as words used "in the enumeration of fauna and flora."

Consequently, Stubbs (*ibid*.) notes that

> There are two closely related key ideas. First, meaning is typically dispersed over several word-forms which habitually co-occur in text. Second, these co-occurring word-forms 'share' semantic features. […] words are said to be 'co-selected', such that words in the collocation are 'delexicalized' (Sinclair 1991: 113; Bublitz 1996). For example, in a phrase such as *physical assault*, the adjective adds little to the meaning of the noun, but merely emphasizes or focuses on an expected feature: the default interpretation of an assault is that it is physical. (In a phrase such as *intellectual assault*, the adjective would have its own independent meaning.) The same phenomenon […] can be seen in common collocations such as *added bonus, advance warning, completely forgot, full circle, general consensus, heavy load*.
> (STUBBS, 2002, p. 63)

3.1.4- Two principles of the organisation of language: the idiom principle and the open-choice principle

The idiom principle has been postulated by Sinclair (1991) after his investigations into the co-occurrence of words in connected discourse. Hunston and Francis (2000, p. 21) note that "the observation that meanings are made in chunks of language that are more-or-less predictable, though not fixed, sequences of morphemes leads Sinclair to an articulation of the 'idiom principle'." Sinclair (1991) describes the idiom principle as follows.

> One of the main principles of the organization of language is that the choice of one word affects the choice of others in its vicinity. Collocation is one of the patterns of mutual choice, and idiom is another. The name given to this principle of organization is the idiom principle. (*Ibid*., p.173).
>
> The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments. To some extent, this may reflect

> the recurrence of similar situations in human affairs; it may illustrate a natural tendency to economy of effort; or it may be motivated in part by the exigencies of real-time conversation. (*Ibid*., p.110).

These 'semi-preconstructed phrases' which reside at the basis of the idiom principle are also discussed by Hunston and Francis (2000, p. 21), who argue that if such phrases "are the general rule in language rather than the exception, they are able to be incorporated into the normal organising principle of language as the idiom principle."

However, Sinclair mentions that, since the idiom principle is not sufficient to account for all instances of language use, people can also resort to the "open choice principle", which is

> […] a way of seeing language text as the result of a very large number of complex choices. At each point where a unit is completed (a word or a phrase or a clause), a large range of choice opens up and the only restraint is grammaticalness.
> This is probably the normal way of seeing and describing language. It is often called a 'slot-and-filler' model, envisaging texts as a series of slots which have to be filled from a lexicon that satisfies local restraints. At each slot, virtually any word can occur. […] virtually all grammars are constructed on the open-choice principle. […]
> It is clear that words do not occur at random in a text, and that the open-choice principle does not provide for substantial enough restraints on consecutive choices. We would not produce normal text simply by operating the open-choice principle. […] (SINCLAIR 1991, p. 109, 110)

Therefore, Sinclair argues that "the principle of idiom is put forward to account for the restraints that are not captured by the open-choice model." (*ibid*, p. 110) and presents some features of the idiom principle, as follows.

> a. Many phrases have an indeterminate extent. As an example, consider *set eyes on*. This seems to attract a pronoun subject, either *never* or a temporal conjunction like *the moment, the first time*, and the word *has* as an auxiliary to *set*. How much of this is integral to the phrase, and how much is in the nature of collocational attraction?
> b. Many phrases allow internal lexical variation. […]
> c. Many phrases allow internal lexical syntactic variation. […]
> d. Many phrases allow some variation in word order. […]
> e. Many uses of words and phrases attract other words in strong collocation; […]
> f. Many uses of words and phrases show a tendency to co-occur with certain grammatical choices. […]
> g. Many uses of words and phrases show a tendency to occur in a certain semantic environment. […]
> (SINCLAIR *ibid*., p. 111-112.)

Most importantly, we have to bear in mind the fact that users interpret utterances through the two principles. As Hunston and Francis (2000, p.22) have noted, "[…] any stretch of language can at any one time be interpreted according to the idiom principle and according to the open-choice principle. Such an option is not, however, open to us."

Sinclair (1991, p. 114) demonstrates this by arguing that

> For normal texts, we can put forward the proposal that the first mode to be applied is the idiom principle, since most of the text will be interpretable by this principle. Whenever there is good reason, the interpretive process switches to the open-choice principle, and quickly back again. Lexical choices which are unexpected in their environment will presumably occasion a switch; choices which, if grammatically interpreted, would be unusual are an affirmation of the operation of the idiom principle.

Hunston (2002) further exemplifies the two options available through the two principles of interpretation by providing the example case of the phrase *grasp the point*. When it is interpreted according to the idiom principle, it means "understand the main idea of something;" if interpreted according to the open-choice principle it means "take hold of the sharp end of something" (HUNSTON, *ibid*., p. 145). She argues that "in practice, however, only one interpretation is activated, and a few readers will be undecided as to which meaning is intended in a sentence such as: Perhaps, finally, this terrible accident will help the islanders grasp the point." (*Ibid*.) According to the author, there is evidence that in ordinary language a phrase overlaps with the next phrase very frequently; therefore, "the typicality of each phrase and the originality of their co-occurrence combine. This has been called 'pattern flow' (HUNSTON and FRANCIS 1999) and 'collocation cascade' (Gledhill 1995)." (HUNSTON, *ibid*.)

In conclusion, there is considerable evidence for the usefulness of the idiom principle in explaining how discourse in English is composed. It accounts not only for fixed phrases but for "much less tangible preferences of phrasing that appear to exist," according to Hunston (*ibid*. p. 147) since it "appears to be a considerably more powerful theory than previous approaches to the phraseology of English." The author notes that

> The idiom principle and the open-choice principle together provide a theoretical account for two observations: that phraseology is extremely pervasive in English, and that phraseology alone cannot account for how sentences or utterances are made up. The principles also account for the fact that ambiguity causes much less of a problem in everyday communication than might be expected from the many words that have more than one meaning. As an explanation of how sentences are interpreted, the theory is extremely persuasive. (HUNSTON, *ibid*. p. 149).

The idiom principle, therefore, can be described as a feature of discourse that is paramount to the construction of its internal organisation. Since the focus of the present research is on the patterns of *yeah* and *yes* responses and their phraseologies, we could argue that the idiom principle operates across both the responses and the initiation moves that originate them and, as a consequence of that, the phraseology of responses and their initiation moves should be regarded as a continuum in spoken discourse. This shall be further discussed in 3.2.3 and in the discussions of the findings (*cf*. Chapter 6).

In Chapter 3.2 we shall present an overview of the importance of very frequent words and phrases in corpus studies. The words yeah and yes, which are the central nodes of the corpus under investigation in the present work, are among the most frequent words in spoken corpora (*cf*. the previous chapters). The insights from word frequency and the usefulness of a word that will be discussed in 3.2 shall illustrate some aspects of language use which frequently go unnoticed in studies into spoken English.

3.2- The importance of very frequent words and phrases to corpus linguistics

3.2.1- The importance of very frequent words and phrases to the evaluation of instances of actual language use

Research in the field of corpus linguistics has invariably dealt with one of its most prominent features, namely the most frequent instances of any word or phrase. Interestingly enough, the most frequent words and phrases have led researchers to develop different approaches and theories to their occurrences since they offer various insights into the language under investigation.

Nation and R. Waring (1997) argue that although a language makes use of a large number of words, not all of these words are equally useful. The authors note that one measure of usefulness is word frequency, that is, "how often the word occurs in normal use of the language." (NATION and WARING *apud* SCHIMITT and MCCARTHY, 1997, p. 8)

The authors exemplify this concept through the case of the word *the*, which, according to them, from the point of view of frequency "is a very useful word in English. It occurs so frequently in English that about 7 per cent of the words on a page of written English and the same proportion of the words in a conversation are repetitions of the word *the*." (NATION AND WARING, *ibid.*)

Observations such as the ones above can be of interest to second language learners and second language teachers since, as Nation and Waring highlight,

> if a learner knows these words, that learner will know a very large proportion of the running words in a written or spoken text. Most of these words are content words and knowing enough of them allows a good degree of comprehension of a text. (NATION AND WARING, *ibid*., p. 9)

However, the analyses of corpora have demonstrated that a small number of words in English occur very frequently. Nation and Waring (*ibid*.) illustrate this by presenting a table

derived from a corpus of written texts. This table shows "what proportion of a text is covered by certain numbers of high frequency words" (NATION AND WARING, *ibid.*):

Table I *Vocabulary size and text coverage in the Brown corpus*

Vocabulary size	Text coverage
1,000	72.o%
2,000	79.7%
3,000	84.0%
4,000	86.8%
5,000	88.7%
6,000	89.9%
15,851	97.8%

(FRANCIS and KUCERA, 1982 *apud* NATION and WARING, *ibid.*)

In their analysis of table I, Nation and Waring argue that knowing only 2000 word families would cover around 80% of written texts and 96% of oral texts. This might sound quite encouraging to language learners since, with a vocabulary of 2000 words, they would know four words out of five. But, according to the authors, this ratio would not allow for reasonably successful guessing of the meaning of the unknown words. "At least 95 per cent coverage is needed for that. Research by Laufer (1988a)[25] suggests that 95 per cent coverage is sufficient to allow for reasonable comprehension of a text" (NATION and WARING, *ibid.*, p. 10).

Lists of high frequency words, thus, represent a powerful source for course designers and teachers in that they can provide the basis for the vocabulary component of a language course and enable them, according to Nation and Waring," to judge whether a particular word deserves attention or not, and whether a text is suitable for a class" (*ibid.*, p. 17).

Sinclair (1991) reminds readers of the importance of investigating the different senses of very frequent words in large corpora, since words that have more than one meaning, sense, or usage "occur in very uneven distribution" (*ibid.*, p.101) . According to the author, frequent words have, in general, a more complex set of senses than infrequent words; therefore, the "accumulation of instances of a frequent word is not just more of the same, but ever more clear evidence of complexity." Furthermore, since some words are much more common than others, some senses of one word are much more common than other senses of the same word "many times more common." (SINCLAIR, *ibid.*, p.102.)

[25] LAUFER, B. What percentage of text-lexis is essential for comprehension? In: LAURÉN, C.; NORDMANN, M. (Eds.) Special language: from humans thinking to thinking machines. Clevedon: Multilingual matters, 1988 *apud* NATION and WARING, *ibid.*)

Another example of the usefulness of information about the frequency of words and phrases resides in its contributions to field of lexicography. As described by Hunston,

> Because a corpus can show the diversity of use, and the importance, of very frequent words, current learner's dictionaries tend to include more detailed information than the old ones did about these words. A rough indication of this can be given by comparing the number of senses given for certain frequent words in different dictionaries. […]
> Many of the increases in the number of senses is explained by more information being given about the very frequent uses […].
> There are also 'new' senses, that is, meanings or uses that seemed unimportant before a corpus showed how frequent they were. […]
> Another innovation in dictionaries that has been made possible by the use of a corpus is the inclusion of explicit frequency information. (HUNSTON, 2002, p.97)

It is interesting to note that Sinclair's descriptions of the changes introduced by the use of a corpus in the Cobuild dictionary associate its new dimension of providing help in the writing of compositions to the new, more communicative role of the learner and, also, to pragmatics, the structure of the spoken language and its phraseology. In other words, the work of the team of lexicographers under his supervision spanned beyond the more immediate task of dictionary writing to encompass spoken English and its phraseology from the perspective of pragmatics.

Therefore, some of the general observations made by Sinclair about the findings obtained during the development of the Cobuild dictionary (SINCLAIR 1987a) illustrate how the evidence encountered in a large corpus can contribute to various aspects in language studies. According to Sinclair,

> *a* In nearly every case, a structural pattern seemed to be associated with a sense. Despite the broad range of material in the corpus, when instances were sorted into 'senses', a recurrent pattern emerged. […]
> *b* In a large number of cases – including most of the common meanings of the common words – the sense and the phraseology seemed to have a closer relationship than that in *a*. There was not just a typical syntax, but a typical pattern of lexical collocation as well, and a distribution of meaning across a number of words. Instead of the normal assumption that meaning is a property of the word, except in a number of idiomatic phrases, it was clear that in these central patterns of English the meaning was only created by choosing two or more words simultaneously and disposing them according to fairly precise rules of position. The account of some phrasal verbs with *set* […] gives some examples, and to this can be added dozens of phrases like *set fire to, set on fire, set eyes on, set free*.
>
> These observations […] guided our decisions about how to represent grammar in the dictionary. There was in practice no clear distinction between grammar and lexis, and grammatical rules merged with restrictions in particular instances, and those restrictions ranged from the obviously grammatical to the obviously lexical. (SINCLAIR 1987c, p. 109-110)

As already mentioned in this chapter, the contributions of corpus studies to lexicography exist in a parallel with their contributions to applied linguistics. The previous paragraphs have demonstrated some of the ways in which an awareness of very frequent

words and their phraseologies can contribute to language description and to an understanding of the pragmatic dimensions of words and phrases, including their uses in spoken language.

3.2.2- Expanding the theoretical framework of corpus linguistics: patterns in language pedagogy

3.2.2.1- The contributions of corpus studies to language teaching approaches and syllabus design

The focus on very frequent words, phrases and their patterns, which has evolved in corpus studies, especially in the descriptions provided by Sinclair, has, among other contributions, led to the creation of new approaches to language teaching and produced some different views of syllabus design. This can be best exemplified through the cases of the lexical approach and the lexical syllabus.

The development of the lexical syllabus can be mentioned as a good example of how practice has influenced theory, which, in its turn, has led to practical applications. However, it should be noted that the cases of the lexical approach and the lexical syllabus have been included in this chapter in order to present some other perspectives of the theoretical framework involved in corpus analyses and studies into phraseology. Our aim in this section is not to argue in favour of the lexical approach, since that would require a comprehensive study of its applications. Our aim is, rather, to proceed in discussing the developments and implications of the theoretical aspects involved in corpus linguistics. Therefore, it is the theoretical background and the findings presented in the current literature about the lexical approach and the lexical syllabus that are of interest for the purposes of this section.

The description of the origins of the lexical syllabus presented by Renouf (1987) serves to illustrate how the awareness of the importance of very frequent words and phrases has led some researchers to re-think some former practices and propose a new approach to syllabus design. According to Renouf (1987), lexis had been a neglected area in applied linguistics, and particularly in language-teaching, over many years. The power of tradition was responsible for that because language was represented a series of syntactic structures, until more recently, when it came to be viewed as series of socially-motivated utterances. It had never been presented as a network of lexis and lexical relations.

Therefore, the author notes that the use of a computerised corpus

> […] would accelerate the search for relevant data on each word, allow us to be selective or exhaustive in our investigation, and supplement our human observations with a variety of automatically retrieved information.
> We planned that the resultant analysis would, in the first instance, form the basis for a lexical syllabus for the proposed new English Course […] to be written by Jane and Dave Willis. […](RENOUF 1987, p. 168, 189)

The central concept of the lexical syllabus is described by Hunston (2002, p. 155), who mentions that

> The notion of a 'lexical syllabus' was proposed in a paper by Sinclair and Renouf (1988), and finds its fullest exposition in D. Willis (1990). The term is occasionally (mis-)used to indicate a syllabus consisting only of vocabulary items, but as Sinclair, Renouf and Willis use the term, it comprises all aspects of language, differing from a conventional syllabus only in that the central concept of organisation is lexis.

As a writer of the English course devised by the Cobuild project, Willis (1990) argues for methodologies which "[…] maximise meaningful exposure to and use of language […]" (WILLIS *ibid*., p. iv). The type of exposure that he proposes includes learners' understanding and production of graded language (in order to avoid demotivation), exposure to the commonest patterns and meanings in the language (which, he argues, are those they are most likely to meet when they begin to use language outside the classroom), and an itemising of the language syllabus which would "[…] expose students to language, but also to highlight important features of learners' language experience and to point to what language we might […] expect them to have learned." (WILLIS *ibid*.)

According to Sinclair (*apud* Willis, *ibid*.), the computational analysis provided by the Cobuild project should provide the basis for the lexical syllabus of a new coursebook. The underlying argument of Sinclair's proposal in favour of a lexical syllabus, in Willis's words,

> […] was to do with utility and with the power of the most frequent words of English.
> The most frequent words of English account for around 70% of all English text. That is to say around 70% of the English we speak and hear, read and write is made up of the 700 commonest words in the language. The most frequent 1,500 words account for around 76% of text and the most frequent 2.500 for 80%. Given this, we decided that word frequency would determine the contents of our course. (WILLIS, 1990, p. v.)

The main features of the lexical syllabus can be briefly summarised as including a focus on the commonest patterns in their natural environment, *i.e.*, it emphasises the importance of natural language, the creation of a learners' corpus, and the encouragement of learners to examine that corpus and generalise from it.

Hunston (2002, p. 189) presents an interesting description of the processes involved in the design of a lexical syllabus. She enumerates a series of elements that are distinctive features of the lexical syllabus and cites some examples.

> At its most simple, the argument is that it makes sense to teach the most frequent words in a language first. Sinclair and Renouf argue that 'the main focus of study should be on (a) the commonest word forms in the language; (b) the central patterns of usage; (c) the combinations which they usually form' (1988:148). Their point is that the most frequent words have a variety of uses, so that learners acquire a flexibility of language fairly easily. In addition, the main uses of the most frequent words cover the main points of grammar, if in an unfamiliar form.

Sinclair and Renouf (1988, p. 155 *apud* HUNSTON, *ibid.*) quote the example case of the verb *make. Make* is word with many uses, some of which are rarely covered in most beginners' courses. The authors verified that the most frequently occurring use of *make* is in combinations such as *make decisions*, *make discoveries*, *make arrangements*, "rather than in the more concrete *make a cake*, etc". The authors thus mention that *make* is used more frequently as a "delexical verb than as an ordinary verb". They complement their findings by arguing that "an English course that focuses only on the concrete sense of *make* denies the learner the opportunity to express sophisticated meanings with a simple verb."

Another example of a very frequent word with multiple uses is *back*. According to Sinclair (1999, *apud* HUNSTON *ibid.*), "it is 95th in frequency in the Bank of English, ahead of, for example, *get, may, how, think, even* and *us"*. The reason for this frequency is that it is used in phrases such as get the *bus back, come/go back, look back, move back, turn back*, as well as a noun: *behind your back, at the back* (SINCLAIR, ibid.). The author thus advocates that "teaching the typical uses of *back* introduces the learner to a large amount of language though not a massive vocabulary."

Sinclair and Renouf's (*ibid.*) remarks in relation to the extent and quality of the vocabulary component in the lexical syllabus is noteworthy. They mention that

> almost paradoxically, the lexical syllabus does not encourage the piecemeal acquisition of a large vocabulary, especially initially. Instead, it concentrates on making full use of words that the learner already has, at any particular stage. It teaches that there is far more general utility in the recombination of known elements than in the addition of less easily usable items.

In relation to grammar in a lexical syllabus, Sinclair and Renouf argue that a separate listing of grammatical items is unnecessary:

> If the analysis of the words and phrases has been done correctly, then all the relevant grammar etc should appear in a proper proportion. Verb tenses, for example, which are often the main organizing feature of a course, are combinations of some of the commonest words in the language. (SINCLAIR and RENOUF, 1988, p. 155 *apud* HUNSTON 2002., p.190.)

In this respect, Willis (1990, p. 15) argues that the frequent lack of balance in the treatment of grammar materials and the number of patterns presented in ELT provide a very restricted picture of the grammar of English. Willis exemplifies this through mention of certain grammatical features, such as the passive, the second conditional, and reported statements, which are traditionally treated in coursebooks as difficult and whose presentation is often delayed until the intermediate level. In a lexical syllabus, these features would be taught through an emphasis on the meanings of the words which are central to the patterns of such features, such as the participle, 'would' and the noun phrase, respectively. Therefore, rather than being regarded as difficult structures, they would be treated as words and phrases whose understanding by learners become productive features. It should be noted that what Willis refers to as "structures that are traditionally considered difficult" are grammatical structures, which are different from the structures of patterns which he refers to in other parts of the book, as well as Sinclair does in his work (*cf*. Chapter 3.1).

In such circumstances, a lexically based approach would, firstly, offer "powerful generalisations" (for example, "once learners are aware of the potential of the past tense and *would* to encode hypothesis, they are in principle capable of producing [adequate utterances]"; WILLIS, *ibid*., p. 23). Secondly, in order to foster powerful generalisations from learners, the lexical approach provides more evidence "on which to base useful generalisations about the language" (WILLIS, *ibid.*).

In conclusion, a focus on words provides, according to Willis,

> the raw material to make more powerful generalisations [and] seems to offer learners the potential to create structures for themselves. Word forms are also easily recognisable and easily retrievable. This is not always the case with structures. […] If we are to adopt a strategy which aims at awareness raising, therefore, there are good arguments for highlighting meaning; and if we are to do this, the most effective unit is likely to be the word rather than the structure.

Such accurate perception of the linguistic form is also described by Lewis (1996) as essential in a lexical approach. The author argues that a lexical approach "advocates a total re-evaluation of the language which is offered to students, and how that language is analysed." In this respect, collocations, multi-word items and patterns in general require from teachers a new way of looking at language, otherwise "they cannot be expected to help students to observe the language to which they are exposed in ways which are maximally useful for them." (LEWIS, 1996, p.14). This emphasis on chunks in language use involves, in the case of written English, an awareness of words, word partnerships, and sentence heads and frames, the latter two typical of academic English, according to Lewis (*ibid*.). In spoken English, it involves "identifying the (pragmatic) meaning of whole utterances, many of which are far

from transparent" (*ibid*.). Furthermore, the importance of these utterances as "a major new emphasis within a lexical perspective" is corroborated by knowledge that

> students can usefully employ a repertoire of at least several hundred, if not many thousand, institutionalized expressions[26]. Such expressions are central to effective spoken communication, both receptive and productive. These highly probable sentences, as well as forming part of dialogues for explicit study, should feature frequently in all language practice materials […] [in a] lexical perspective. (LEWIS, 1996, p.15.)

The presentation of frequently occurring sequences in a corpus syllabus is considered by Hunston (2002, p. 190) as a good alternative to a "word-by-word account" of frequencies (since this would be very lengthy). The author notes the importance of sequences in a lexical approach and cites the example of studies into prefabs[27] in native speaker and learner corpora which were carried by De Cock et al (1988) and De Cock (1998). The aim of their studies was to test the hypothesis that learners tend not to use formulae as frequently as native speakers do.

Another important study into sequences is found in Biber *et al* (1999), who demonstrate that "three-word bundles[28] are much more frequent than four-word", both kinds of bundles are much more frequent in conversation than in academic prose and that in conversation the bundles comprise more of the total word-count (28%) than they do in academic prose (20%)" (*cf*. HUNSTON, *ibid*., p. 191).

Findings of the types mentioned in the preceding paragraphs demonstrate that the changes introduced in syllabus design and the new approaches to language teaching are constantly supported by the evidence and the theoretical developments obtained from corpus studies. They also indicate that the field of corpus linguistics, which is characterised by its

---

[26] Lewis (1996, p. 10) provides only the following examples of "institutionalized utterances or fixed expressions": "*I'll see what I can do*", and "*It's not the sort of thing you think will ever happen to you*." Since the nomenclature in this field varies, we will not present any additional definitions of the term in this chapter, since a brief discussion shall be presented in chapter 4.

[27] De Cock et al (1998, p. 79) refer to "prefabs" as "prefabricated or prepatterned expressions" and "formulae or formulaic expressions, *i.e.* frequently used multi-word units that perform pragmatic or discourse structuring functions." They note that their analyses were restricted to continuous prefabs, *i.e.*, unbroken sequences of words, due to the fact that they used a programme for automatic extraction which only allowed for the investigation of continuous prefabs.

[28] Biber et al (1999, p. 990) define "lexical bundles" as recurrent expressions, regardless of their idiomaticity, and regardless of their structural status. That is, lexical bundles are simply sequences of word forms that commonly go together in natural discourse. [...] A lexical bundle is defined here as a recurring sequence of three or more words. Shorter bundles are often incorporated into more than one longer lexical bundle. For example, the three-word lexical bundle *I don't think* is used in many four-word bundles, such as *but I don't think* , *well I don't think*, *I don't think so*, and *I don't think I*. [..] A combination of words must recur frequently in order to be considered a lexical bundle. In most cases, these bundles are not structural units, and they are not expressions that speakers would recognize as idioms or other fixed lexical expressions.

capacity of providing data that are constantly updated, has introduced changes into language pedagogy that will require the constant feedback and support from theory.

The next section will present a brief discussion of the rationale behind the production of discourse-based grammar books.

3.2.2.2- Corpus linguistics and the development of grammar books

Word frequency and patterns, again, have had a decisive role in the re-shaping of grammar books. In this section, we will present a brief overview of some of the most representative cases.

One of the first grammars developed in this perspective is the *Cobuild grammar* (SINCLAIR *et al*, 1990). According to Hunston (2002, p. 99), the Cobuild grammar was elaborated in accordance with one of the two existing approaches to dealing with frequency in grammar books. That is, it is based on a focus on usages that are relatively frequent. The Cobuild grammar thus presents "the main patterns of English", which are described, and "the typical words and phrases found in each pattern are listed" (SINCLAIR *et al*, 1990, p. v.). It contains many lists (of nouns, adjectives and others), which indicate that "words with similar behaviours tend to have similar meanings" (HUNSTON 2002, p. 104).

Another approach to grammar, according to Hunston (*ibid*., p. 99), consists in their presentation of statistical figures of the frequencies of the different usages of words and phrases. Some writers of grammar books give precise statistical information based on frequency counts in specially designed and annotated corpora. The author mentions the case of *An empirical grammar of the English verb system* (MINDT, 2000), which deals with the distribution of meanings across a given form.

> Mindt (2000: 224), for example, identifies four meanings of the present perfect (the indefinite past, past continuing into present, the recent past, and a use indicating that an action is completed, though not at an unspecified time). Of these, the first (indefinite past) accounts for almost 80% of all occurrences of the present perfect, with the second (past continuing into present) accounting for all but 5% of the others. […] This is in conflict with many coursebooks which teach uses such as I have lived here for 12 years (past-into-present) or they have recently had their child (recent past) as prototypical, when in fact they are less common than the indefinite past use. Most of the other tense forms presented by Mindt have similarly asymmetrical patterns of use. […]
> This kind of work is probably most useful when frequency can be linked to discourse. (HUNSTON 2002, p. 99)

However, Hunston notes that the limitations of frequency information in classroom materials can sometimes suggest that very infrequent uses can legitimately be ignored. The

author argues that very infrequent uses found in corpora cannot be ignored by syllabus designers and coursebook writers.

Another recent application of corpus studies to grammars can be found in spoken discourse grammars. For example, McCarthy (2001) mentions work by Hughes and McCarthy (1988), who present several examples of grammatical features derived from a corpus which require a new perspective when examined in their actual contexts, both spoken and written. Although they use corpus evidence, "their approach goes beyond statements of statistical distribution of items […] to qualitative interpretations of grammatical relations based on evidence across a range of texts […]." (McCARTHY 2001, p.107.) The author also notes that the evidence demonstrates the need for are-assessment of the sentence as a viable unit of grammatical description. Since "well-formed sentences are the exception rather than the norm in many kinds of everyday conversation (*e.g*. causal talk, some service encounters), […] the clause emerges as a better candidate for the base unit of description."

Another very important aspect approached by McCarthy is the fact that the units of grammar are often co-created by participants. Therefore, an element of one speaker's turn may only be grammatically coherent when seen as a continuation of another speaker's utterance. This, according to the author, illustrates that

> grammar as joint-construction, rather than just an encoding by one speaker and a decoding by another, and one is reminded of Farr and Rommetveit's (1995:265) admonition that 'when expression is […] equated with "encoding" and impression with "decoding" […] one has bought the language of the telecommunication'. In their view, grammar would partake in that 'commonality' that is the hallmark of the discourse process, and 'commonality is established when two persons construct a temporarily shared world by engaging in dialogue' (*ibid*.:271). The shared world is as much expressed in grammar as it is in lexical selection, and co-construction is one of its key manifestations.

At this point, we could mention the correlation between McCarthy's views of grammar as 'joint-construction' and the shared world as constituted by grammar and lexical selection with our view of the phraseology of responses and their initiation moves as a continuum in spoken discourse (*cf*. Chapter 3.1.3). As already mentioned, the phraseologies of responses operate in connecting discourse within the speaker's move and in relation to the interlocutor's initiation move.

Furthermore, the new perspective of discourse grammar books allows for the development of further theoretical applications of the insights they offer, thus introducing additional changes.

> Beyond-the-sentence investigations of grammatical choices suggest that discourse grammars do more than just add 'bolt-on-extras' to existing sentence grammars, and

> precipitate a complete re-assessment of how grammars are written, especially spoken ones. […]
> Celce-Murcia (1991) sees value in a discourse-based approach to grammar as stemming from a study of learners' communicative needs and the assembly of a corpus of material relevant to those needs; after these stages, and only then, should the decision be taken as to the most useful grammar to be taught. (MCCARTHY 2001, p. 109)

One of the major corpus-based grammar books is the *Longman grammar of spoken and written English (LGSWE*; BIBER *et al*, 1999). This grammar book provides statistical information based on frequency counts and the grammatical features are presented in terms of a comparison of frequency between four registers, namely conversation, fiction, newspaper language and academic prose. The *LGSWE* adopts a corpus-based approach and the *LGWSE* corpus contains over 40 million words (*cf*. BIBER *et al*, *ibid*.). The authors argue that "although the grammar is primarily organized along structural lines, the descriptions emphasize not only their formal properties but also the use of these structures." (BIBER *et al*, *ibid*., p. 4), which constitutes "an entire extra dimension for grammatical description". Furthermore, it investigates the most important patterns of use "in data-intensive ways" (*ibid*.) and identifies the differences between American English and British English.

The features of the *LGWSE* are discussed by Hunston (2002), who argues that

> Work of this kind raises the question as to whether the registers selected for comparison are sufficiently homogeneous, or whether they themselves are open to the charge of being monolithic. […]
> Another striking feature of the *Longman grammar of spoken and written English* is the degree to which lexical information forms an integral part of the grammatical description. […] In this concern for lexis the writers concur with Sinclair, who, in fact, rejects the distinction between lexis and grammar. […] (Note that Sinclair *et al* take a rather different view of the importance of frequency than do Biber *et al*. Whereas Biber *et al* tend to give the most frequent verbs to be found with a certain form in a given register, Sinclair *et al* list the verbs which are most important to a given meaning made in a particular way.) (HUNSTON, *ibid*., p. 104.)

The descriptions above demonstrate the ways in which the rationale of corpus linguistics and its theoretical framework have been incorporated into the development and metalanguage of discourse-based grammars. The evidence obtained from the analysis of large corpora represents a new paradigm in language description, whose practical applications are already available to the public.

In section 3.2.3 we will discuss the issue of the boundaries between grammar and lexis.

3.2.3- The boundaries between grammar and lexis: grammatical and lexical words, the evidence from corpus studies and the case of spoken English

Grammatical (or function) words and lexical (or content) words (*cf*. McCARTHY, 1991; McCARTHY and CARTER *apud* SCHMITT and McCARTHY, 1997; STUBBS 2002) have been at the centre of the debate about word frequency and usage in corpus analyses. Therefore, some considerations about the current debate have been included in this section.

Vocabulary is divided into the two major categories of content words, *i.e.*, words which tell us what a text is about (noun, adjective, adverb and main verb), and function words, *i.e.*, words which relate content words to each other (auxiliary verb, modal verb, pronoun, preposition, determiner and conjunction). The distinction is made in most grammars of English, but since many linguists make essentially the same distinction, there are several terms in use, according to Stubbs (2000). Content words are also referred to as major, full and lexical words. They carry most of the lexical content, in the sense of being able to make reference outside language. Function words are also referred to as minor, empty, form, structural and grammatical words. They are essential to the grammatical structure of the sentences. Their function is internal to the language, for example, in making explicit the relation of lexical words to each other (*cf*. Stubbs, *ibid*., p. 39). Stubbs notes that it is possible to conceive of a communicative system which has only content words, but not of a system which has only function words. "For example, in a telegram one can omit function words and still have a comprehensible message" (*ibid.*).

However, the boundary between the two word classes is not perfectly clear-cut. For example, modal verbs may express obligation, permission and ability, and therefore convey content; and pronouns have extra-linguistic reference. On the other hand, content and function words have strikingly different formal characteristics, as noted by Stubbs (*ibid.*) since "content classes have many members (there are tens of thousands of nouns, but only a couple of dozen pronouns), and are open to new words (for example, new nouns and verbs are being constantly invented)" (STUBBS *ibid*., p. 39-40).

McCarthy and Carter (1997) approach the issue of content and function words from the perspective of spoken language. They observe that "frequency lists for everyday spoken language differ significantly from those dependent only on written databases" (*ibid*., p. 23)

and present Table I (which is partially reproduced here) in order to demonstrate the differences that arise in the comparison of lists of the most frequent items in spoken and written data.

Table I Fifty most frequent words from 330,000 words of Cambridge International Corpus (CIC) written data and 330,000 words of spoken data (CANCODE)[29]

	Written	Spoken
1	the	the
2	to	I
3	of	you
4	a	and
5	and	to
6	in	it
7	I	a
8	was	**yeah**
9	for	that
10	that	of
11	it	in
12	on	was
13	he	is
14	is	it's
15	with	know
16	you	no
17	but	oh
18	at	so
19	his	but
20	as	on
21	be	they
22	my	well
23	have	what
24	from	**yes**
……..		

(McCARTHY and CARTER, 1997, p. 23-24.)

It should be noted that this table also demonstrates the high frequencies of the words *yeah* and *yes* in the corpus of spoken English (against no instances in the written corpus). The authors argue that lists such as the one above present a similarity of occurrence of basic function words and some interesting differences. The written list is made up of function words (function words here include all non-lexical, *i.e.* non-contentful items, such as pronouns, determiners, prepositions, modal verbs, auxiliary verbs, conjunctions, etc.), but the spoken list includes a number of lexical words such as *know*, *well*, *got*, *think*, *right*.. The function words dominate the top frequencies of both lists, and, indeed, one of the defining criteria of function words is their high frequency (*cf*. McCARTHY and CARTER, *ibid*.).

[29] The CANCODE spoken English corpus is developed at the University of Nottingham, UK, and sponsored by Cambridge University Press. It consists of 'everyday British English conversational text and is a targeted corpus, classified according to genre-headings such as narrative, language-in-action, service encounter, comment-elaboration , (…)' (*cf.* McCARTHY and CARTER. In: SCHMITT and McCARTHY, *ibid.* p. 22).

However, McCarthy and Carter have also noted (*cf*. STUBBS 2002, above) that "there is no absolute cutoff between function words and lexical words of high frequency". They exemplify this through the case of the word *thing*, and argue that "[…] using frequency alone, without other criteria (*e.g*. whether the word in question belongs to an open or closed set) results in a blurred borderline between 'grammar' and 'vocabulary' words […]." (McCARTHY and CARTER, *ibid*.)

Further examinations of Table 1 above have also demonstrated that

> Some of the "lexical" words which intrude into the high-frequency function word list prove to be elements of interpersonal markers (e.g. *you know, I think*) or single-word organisational markers (*well, right*)[30]. […] *Well* occurs approximately nine times more frequently in spoken than in written; in the written sample, it is number 135 in the list, compared with number 22 in the spoken. […] The hedging word *just* ranks at 33 in the spoken; in the written it ranks at 61 and is two and a half times less frequent. (…) But frequency statistics alone do not tell us everything; examination of a **concordance** produced by the computer is even more helpful. (McCARTHY and CARTER, *ibid*., p. 25.)

Therefore, the issue of the boundaries between grammar and lexis which has been discussed by McCarthy and Carter (*ibid*.) has, also, been approached through the use of corpora. Their findings about the most frequent words in the spoken and written modes have led them to advocate the need for the examination of concordances, a subject that shall be further exploited in Chapter 5. They demonstrate the importance of the analysis of concordances for a better understanding of collocation. They present some interesting findings derived from the analysis of concordances for the subject pronoun *I* plus *'ve got* in typical spoken uses. Different meanings of *I've got* were identified, such as 'I have to deal with' and 'I understand you'. "Neither meaning might crop up in formal, written texts; spoken data is likely to be the best source for such uses" (*ibid*., p. 27).

In addition to that, they further exemplify the differences in distribution and usage between the spoken and the written modes through the cases of the words *start* and *begin* : *start* occurs more or less equally in both modes, while *begin* is more frequent in the written mode (119 occurrences of *begin* in the written against a mere 27 in the spoken); and the words *too* and *also*: the former occurs more or less equally in spoken and written discourse, while the latter is more frequent in the written (289 occurrences of *also* in the written against 107 in the spoken). They also note that in the case of *begin*, "the form *beginning* used as a noun

[30] McCarthy and Carter (*ibid*., p. 25) here mention the work of Stenström, who "discusses such words that belong quintessentially to the spoken mode, and offers a useful set of headings for what she generally refers to as **discourse items**, which include apologies, smooth-overs (*e.g*. *never mind*), hedges (*e.g*. *kind of/ sort of*), and a variety of other types unlikely to occur in the written mode."
(*Cf.* STENSTRÖM, A. B. 1990. Lexical items peculiar to spoken discourse. In: SVARTVIK, J. (Ed.) *The London-Lund corpus of spoken English*: 137-75. Lund: Lund University Press.)

occurs 41 times, but in the spoken only 15 times, reflecting the tendency towards nominalisation in the written mode" (*ibid.*).

Also in relation to the "top 50" spoken and written word-forms presented in Table 1, McCarthy and Carter (*ibid.*) note that the identification of how much of the total text in the corpus they cover is an important aspect to be taken into consideration. They also mention some of the problems that arise in comparative studies of spoken and written corpora, noting, among other facts, that

> word lists consisting of single word-forms (…) may hide the fact that that form regularly occurs as an element of a multi-word expression. For example, how many of the 500+ occurrences of *thing* in the CANCODE spoken sample are embedded within the extremely common expression *the thing is* … (meaning 'the problem is/point is …')? How many are in vague expressions such as *things like that*? Only a concordance can properly reveal whether *thing* is occurring in this way or not. (*Ibid*., p. 29)

Thus, the observations made by McCarthy and Carter (*ibid.*) also reflect the insights found in the work of Sinclair, which has already been described in this chapter. In the same manner that Sinclair has expressed it, McCarthy and Carter present their findings (also obtained from corpora) to conclude that the frequent words they encountered in their analyses may be parts of expressions, or phrases, therefore demonstrating that the boundaries between grammar and lexis are not clear. They also argue in favour of corpus studies of spoken texts "[…] for it is in the study of spoken texts that significant differences have emerged which prompt a reassessment of some aspects of vocabulary pedagogy" (McCARTHY and CARTER *ibid*., p. 20).

Therefore, the present study, with its focus on *yeah* and *yes* responses, includes an investigation of the phraseologies that co-occur with the words *yeah* and *yes*, which are very frequent lexical words in spoken English. Furthermore, since these two very frequent words are used in spoken responses, it could be argued that the phraseologies that co-occur with them perform the role of connecting discourse across different speakers' utterances, in addition to connecting it between the same speaker's utterances. Thus, an investigation into how these words and their senses are currently defined and described in dictionaries (especially those based on corpora) will be presented in section 3.2.4, in an attempt to illustrate some of the latest changes that have been introduced by the use of corpora.

3.2.4- Some definitions and descriptions of the senses of two very frequent words: *yeah* and *yes*

If we keep in mind the fact that YYRs (which are the focus of the present study), as instances of affirmative responses, are phraseologies that operate in connecting discourse, it is important to remember John Sinclair's (1991) remarks about the most frequent words in English (*cf*. section 3.1 in the present study). The word *yes* appears at the 81st position in the *Bank of English* corpus, which is considered a very high frequency (*cf*. SINCLAIR 1991). To Sinclair's remarks, we could add the figures presented by McCarthy and Carter (1991), which are further discussed in Section 3.2.3 below. Here the authors demonstrate that *yeah* and *yes* appear at high frequencies in the Nottingham corpus, too.

Although the figures presented by Sinclair and McCarthy and Carter include uses of *yeah* and *yes* alone with various illocutionary values (which are not identified in their work and may include instances of *yes* alone), we know that *yeah* and *yes* are very frequently used in instances of longer responses (*cf*. Silveira Pedro, 1999, and Chapters 5 and 6 in the present study), which very often connect the contents of the response to additional information provided by the same speaker, for example (this and other types of sequences shall be described in Chapter 5).

Therefore, the different senses of *yeah* and *yes* that can be found in modern dictionaries shall be of interest to the present study, although it does not constitute a work in lexicography. As the discussions in 3.2.1 have demonstrated, the insights obtained from computerised corpora have had useful applications in lexicography and applied linguistics in general. Furthermore, the work of some lexicographers who have worked with corpora has made available the new evidence of the different senses of a word and their phraseologies, representing, it could be argued, the most immediately available resources to language users.

The definitions of *yeah* and *yes* in some currently available online versions of well-known dictionaries are presented below. All these dictionaries are available in the internet, except for the *Collins Cobuild English language dictionary* (*CCELD*) and the *Oxford advanced learner's dictionary of English* (*OALDCE*). The decision for presenting a greater number of samples of online versions of dictionaries was due to the fact that these versions are constantly updated on the internet. The *CCELD* is only available in printed format, and the

version of the *OALDCE* presented was printed in 1987; this version was included here in order to contrast an earlier edition with some newer, corpus-based editions of similar dictionaries.

The criteria that has been adopted for the order of the presentation of the dictionary entries is the number of the different senses of the words *yes* and *yeah* (other than noun or adjective) that they contain. They are presented in decreasing order of number of senses, as follows.

1- The *Oxford advanced learner's dictionary online* (*OALDO*; 7th edition; based on the British National Corpus)

yes *exclamation*, *noun*
exclamation
1 used to answer a question and say that sth is correct or true: *'Is this your car?' 'Yes, it is.'* ◇*'Are you coming? Yes or no?'*
2 used to show that you agree with what has been said: *'I enjoyed her latest novel.' 'Yes, me too.'* ◇*'It's an excellent hotel.' '**Yes, but** (= I don't completely agree) it's too expensive.'*
3 used to disagree with sth negative that sb has just said: *'I've never met her before.' 'Yes, you have.'*
4 used to agree to a request or to give permission: *'Dad, can I borrow the car?' 'Yes, but be careful.'* ◇*We're hoping that they will **say yes to** our proposals.*
5 used to accept an offer or invitation: *'Would you like a drink?' '**Yes, please / thanks**.'*
6 used for asking sb what they want: *Yes? How can I help you?*
7 used for replying politely when sb calls you: *'Waiter!' 'Yes, sir?'*
8 used to show that you have just remembered sth: *Where did I put the keys? Oh, yes—in my pocket!*
9 used to encourage sb to continue speaking: *'I'm going to Paris this weekend.' 'Yes ... '*
10 used to show that you do not believe what sb has said: *'Sorry I'm late—the bus didn't come.' 'Oh yes?'*
11 used to emphasize what you have just said: *Mrs Smith has just won £2 million—yes!—£2 million!*
12 used to show that you are excited or extremely pleased about sth that you have done or sth that has happened: *'They've scored another goal.' 'Yes!!'*
13 yes, yes used to show that you are impatient or irritated about sth: *'Hurry up—it's late.' 'Yes, yes—I'm coming.'*
IDM **yes and no** used when you cannot give a clear answer to a question: *'Are you enjoying it?' 'Yes and no.'*
■***noun*** (*pl*. **yesses** or **yeses** /jesɪz/) an answer that shows that you agree with an idea, a statement, etc.; a person who says 'yes': *I need a simple yes or no to my questions.* ◇*There will be two ballot boxes—one for yesses and one for noes.* ◇*I'll put you down as a yes.*
yeah *exclamation* (*informal*) yes
IDM **oh yeah?** used when you are commenting on what sb has just said: *'We're off to France soon.' 'Oh yeah? When's that?'* ◇*'I'm going to be rich one day.' 'Oh yeah?' (= I don't believe you.)*
IDM **yeah, right** used to say that you do not believe what sb has just said, disagree with it, or are not interested in it: *'You'll be fine.' 'Yeah, right.'*

2- The *Collins Cobuild English language dictionary* (*CCELD*; Sinclair, 1987a; based on the Bank of English corpus)

> **Yes, yeses**; **yeah** is an informal word form. **Yes** and **yeah** are used in speech to express different sorts of responses and reactions, and this entry shows the commonest ones. **1** You use **yes**, mainly in spoken English, **1.1** to answer a question to which the answer could be 'yes' or 'no'. Eg *'Did you enjoy it?'_"Yes.'*…[…]. **1.2** to accept an offer. Eg […] '*Do you want a cushion?'_'Oh yes. thank you.*' **1.3** to say that you are willing to do or allow what someone has asked you. Eg '*Will you let us know about it?'_Yes I will''*. **1.4** to tell someone that they have answered a question correctly. Eg […] *'Thirty kilohertz.'_'Yeah that's right*.' **1.5** to show that you are ready or willing to speak to the person who wants to speak to you, for example, when you are answering a telephone or doorbell. Eg. '*Sir* ?'_'*Yes* ?' […].
> **2** You say **yes and no** in reply to a question when you cannot give a definite answer because there are several different possible answers which are partly true. Eg *'Do they actually use computers to diagnose illnesses?'_'Yes.'*
> **3** A **yes** is a person who has answered 'yes' to a question or who has voted in favour of something; also used to refer to their answer or vote. Eg *There were seventeen yeses and only two don't knows*.
> **4** You also use **yes** in a conversation **4.1** to indicate your involvement in the conversation and to say that you agree with, accept or understand what the previous speaker has said. Eg […] *'You'll have to fill in a form when you come.'_'Oh yes, that'll be no problem.*' […]. **4.2** to encourage someone to continue speaking. Eg […]. **4.3** as a polite way of introducing an objection to what the previous speaker has just said. Eg 'It's a waste of time.'_*'Yes but there's still some point in going*.' […]. **4.4** to say that a negative statement or question that the previous speaker has made is wrong or untrue. Eg '[…] *Don't you know Latin*?'_*'Yes, of course I do*.' **4.5** to suggest that you do not believe what the previous speaker has said, especially when you are feeling annoyed or scornful. Eg 'this is a musical cat.'_'Oh, yes?' **4.6** to introduce an opinion, statement or remark. Eg 'What do you think, Jack?'_'Yes, well I agree too.' […]. **4.7** to introduce something that you had forgotten to say and have just remembered. Eg […] *What was I going to mention? Ah yes, accidents*. **4.8** to emphasize and confirm a statement that you are making. Eg […] '*Do you know he crashed the car?'_'No.'_'Oh yeah*.'

(Note: all the senses, except number 3, are identified as *conventions* of the language, in an extra column located to the right side of the entries; number 3 is identified as a countable noun. The information in this extra column, as described in the *CCELD* [Sinclair 1987a, p. xi], consists of grammar notes and semantic relationships.)

3- The *Longman dictionary of contemporary English online* (*LDCEO*; based on the Longman corpus)

> **yes 1** ***adverb***
> W1- Top 1000 written words / S1- Top 1000 spoken words
> **yes** *spoken*
> **1** ANSWER TO QUESTION/STATEMENT
> **a)** used as an answer to say that something is true or that you agree [? no]:
> *'Is that real gold?' 'Yes.'*
> *'It was a great show.' 'Yes, it was.'*
> **b)** used as an answer to a question or statement containing a negative, to say that the opposite is true:
> *'Sarah isn't very intelligent, is she?' 'Yes, she is* (=in fact, she is intelligent)*!'*
> *'There isn't any cereal left.' 'Yes, there is - it's in the cupboard.'*
> **2** ANSWER TO OFFER/INVITATION used as an answer to say that you want something or want to do something [? no]:

'Would you like a sandwich?' ' ***Yes, please****.'*
'Would you like to come with us?' 'Yes, I'd love to.'
3 ANSWER TO REQUEST used as an answer to say that you will do something, or that someone may do or have something [? no]:
'Can I have a glass of water?' 'Yes, of course.'
He proposed to me and I ***said yes****.*
4 yes, but … used to show that you agree with what someone has said, but there is another fact to consider:
'There are still a lot of problems with Jeff's proposal.' 'Yes, but it's the best one we have.'
5 READY TO LISTEN/TALK used to show that you have heard someone or are ready to speak to someone:
'Mike?' 'Yes?'
Yes sir, how can I help you?
6 LISTENING used to show that you are listening to someone and want them to continue:
'And so I tried phoning him ...' 'Yes ...'
7 EXCITED/HAPPY used to show that you are very excited or happy about something:
Yes! Rivaldo's scored again!
8 oh yes
a) used to show that you do not believe what someone is saying:
'There's nothing going on between me and Jane. We're just good friends.' 'Oh yes?'
b) used to show that you have remembered something:
Where's my umbrella? Oh yes - I left it in the car.
9 EMPHASIS used to emphasize that you mean what you have just said, even though it is surprising:
It took ten years - yes, ten whole years - to complete.
Yes, you heard me correctly - I said 1921.
10 yes, yes used to show annoyance when someone is talking to you and you do not want to listen:
'And don't forget to lock the door!' 'Yes, yes, OK.'
11 yes and no used to show that there is not one clear answer to a question:
'Were you surprised?' 'Well, yes and no. I knew they were planning something, but I wasn't sure what.'
? YEAH
yes 2 ***noun***
yes *plural* **yeses** *or* **yesses** [countable]
a vote, voter, or reply that agrees with an idea, plan, law etc:
According to the latest opinion poll, the noes have 60%, and the yeses have 40%.
—yes *adjective*:
a yes vote
yeah ***adverb***
yeah *spoken informal*

4- The *Cambridge advanced learner's dictionary online* (*CALDO*; based on The Cambridge International Corpus)

yes adverb (INFORMAL **yeah** or **yep** or **yah**)
1 used to express acceptance, willingness or agreement:
"Would you like a glass of wine?" "Yes please."
"Do you like Indian food?" "Yes, I love it."
"He's a really nice guy." "Yes he is."
"Report to me at nine o'clock tomorrow morning." "Yes, sir."
"Have you had enough to eat?" "Yes, thank you."
If you'd ***say*** *yes* (= agree) *to the request you'd save a lot of trouble.*
2 used to show that you are listening to someone, or that you are ready to listen and to give them an answer or information:
"Dad." "Yes, what do you want, honey?"

Yes, can I help you?
3 used when you are disagreeing with a negative statement:
"I'm not a very good cook though." "Yes you are - you make wonderful food!"
yes noun [C]
a vote supporting a particular plan of action or an acceptance of an invitation:
"Have you had any replies yet?" "Six yeses and two noes so far."
oh yes (INFORMAL **oh yeah**)
used when you have just remembered something that you were saying:
What was I talking about - oh yes, I was telling you what happened at the party.
yes and no
used when you cannot give a particular answer to a question:
"Is the job going okay?" "Well, yes and no."
yes-man noun [C] DISAPPROVING
a person who agrees with everything their employer, leader, etc. says in order to please them

5- *ENCARTA dictionary*

yes
adverb, interjection
Definition:
1. assent indicator: used, especially in speech, to indicate assent, agreement, or affirmation
- "Do you like ice cream?" "Yes, I do."
- 97 percent of respondents answered yes.

2. indicates contradiction: used to indicate contradiction in response to a negative proposition
- "He won't believe you." "Oh yes he will."

3. mark of attention: used to indicate that somebody is ready to give his or her attention to somebody who has asked for it
- "Doctor?" "Yes?"

4. acceptance: used to accept an offer or a request
- "Would you like some tea?" "Yes, please."

noun (plural yes·es or yes·ses)
Definition:
1. affirmative response: an affirmative response to a question
- Was that a yes or a no?

2. affirmative voter: somebody who votes in the affirmative
- The yeses have 65 percent and the noes 35 percent, so the motion is carried.

interjection
Definition:
exclamation of jubilation: used as a loud exclamation to express triumph, jubilation, or extreme excitement and pleasure (informal)
[Old English gese < gea (see yea) + sie "may it be (so)," form of the verb to be]
say yes to express agreement or consent
Our team won the championship--yes!!!

6- *WIKIPEDIA dictionary*

Adverb
yes
A word used to show agreement or acceptance.
Yes, you are correct.
Yes, you may go to the bathroom now.
A word used to indicate disagreement or dissent in reply to a negative statement.
Usage note
In Old and Middle English, yes was a more forceful affirmative than yea.

> An example of yes used to disagree with a statement: the questions "You don't want it, do you?" and "Don't you want it?" are answered by "yes" if the respondent does want the item, and no if he or she does not. Many languages use a different word for this purpose.
>
> Col Interjection
> yes!
> Used to express pleasure, joy, or great excitement.
> Noun
> yes (plural yeses)
> An affirmative expression; an answer that shows agreement or acceptance.
> Was that a yes?
> A vote of support or in favor/favour of something.
> The workers voted on whether to strike, and there were thirty "yeses" and one "no".
> Verb
> yes
> (transitive) To agree with, to affirm.
> Did he yes the veto?

7- The *Merriam-Webster online dictionary* (*MWOD*)

> Function: *adverb*
> Etymology: Middle English, from Old English *gEse*
> **1** -- used as a function word to express assent or agreement <are you ready? *Yes,* I am>
> **2** -- used as a function word usually to introduce correction or contradiction of a negative assertion or direction <don't say that! *Yes,* I will>
> **3** -- used as a function word to introduce a more emphatic or explicit phrase
> **4** -- used as a function word to indicate uncertainty or polite interest or attentiveness

8- The *Collins dictionary online* (*CDO*)

> Yes
> sentence substitute
> 1. used to express acknowledgment, affirmation, consent, agreement, or approval or to answer when one is addressed
> 2. used, often with interrogative intonation, to signal someone to speak or keep speaking, enter a room, or do something
> 3. noun an answer or vote of yes
> 4. (often plural) a person who votes in the affirmative
> Compare: no 1: History: Old English gese, from ia sie may it be; see yea.

9- The *Oxford advanced learner's dictionary of current English* (*OALDCE*; HORNBY, 1987)

> **Yes** *particle* (contrasted with *no*) expressing agreement, affirmation, consent, etc; 'Can you read this?' '*Yes.*' (Note that *yes* is used as an answer to an interrogative-negative if the complete answer is affirmative: '*Don't you like it?*' '*_Yes*' (*='yes, I do like it'*) ' […] *'Waiter!' _'Yes, sir,'* (*=What do you want, sir*?')
> **-** n [C] affirmation; acceptance: *Answer with a plain 'Yes' or 'No'.*

The definitions and descriptions reproduced above demonstrate that those dictionaries which used computerised corpora as a basis, namely, the *Cambridge advanced learner's dictionary online* (*CALDO*), the *Collins Cobuild English language dictionary* (*CCELD*), the

Collins dictionary online (*CDO*), the *Longman dictionary of contemporary English online* (*LDCEO*), the *Oxford advanced learner's dictionary online* (*OALDO*), present senses of the words *yes* and *yeah* which have not traditionally been included in dictionaries. This contrast can be exemplified through a comparison of the senses encountered in these dictionaries with the senses encountered in the 1987 edition of the *Oxford advanced learner's dictionary of current English* (*OALDCE*; HORNBY, 1987). The *OALDCE* presents a much more limited number of senses (except for the case of the *CDO*, which, due to commercial reasons, presents very few senses, different, though, from the senses found in the *OALDCE* and some of the other dictionaries). A comparison of the number of senses (other than noun, verb or adjective) for *yes* in both the 2006 edition of the *OALDO* and the 1987 edition of the *OALDCE* demonstrates that the former contains thirteen senses and three separate uses in idiomatic expressions, totalling sixteen senses used in responses (in addition to the uses of *yes* and *yeah* as nouns) whereas the *OALDCE* (1987) contains three senses of *yes* and *yeah*, of which one is not stated but demonstrated through an example sentence.

The dictionary entries reproduced above demonstrate that the *OALDO* contains the largest number of senses of *yes* (other than noun or adjective). It is followed by the *CCELD*, *LDCEO* and *CALDO*. The *OALDO*, *CCELD* and *LDCEO* can be described as containing the largest numbers of senses of *yes and yeah*. They either make direct reference to their use in conversation and speech (as in the case of the *CCELD*) or present example sentences in direct speech (*cf*. the *CALDO*, *LDCEO* and the *OALDO*). The *CCELD* presents examples in direct speech too. A distinctive feature of the *LDCEO* is the fact that it includes information about the frequency of *yes* in their corpus. It mentions that *yes* is among the 1000 most frequent written words and 1000 most frequent spoken words.

Some of these dictionaries highlight the use of *yes* in expressions, such as *yes but*, *oh yes, yes and no, oh, yes?* and *yes, right*. In this respect, the *OALDO* introduces *yes and no, oh yeah* and *yeah, right* as idiomatic expressions (*cf*. the abbreviation *IDM* in their equivalent entries), which, once again, demonstrates the inclusion of expressions with *yes* which traditionally did not appear in dictionaries.

Yeah has been described in the entries as the informal equivalent to *yes*. *Yes* has been classified in some of the dictionaries under investigation as an adverb, a noun, an adjective, an interjection, a verb and a sentence substitute. The *CCELD*, however, does not use this system of classification for the word yes, since it aims at identifying the patterns and phraseologies of words. The *CCELD*, however, describes some other words as verbs, nouns, and others in a separate column located to the right of the entries. In this column, *yes* was identified as

'context' (the column could not be reproduced here due to the lay-out of the citation; it should be noted, however, that, in the case of *yes*, there was no word other than 'context' in this column). The classification of *yes* as a 'sentence substitute' is used in the *CDO* only. In addition to that, the *CDO* classifies it as a noun as well. The *CDO* is based on the same corpus as the *CCELD*; although it is an online dictionary, the *CDO* is very economical in its descriptions and definitions (probably due to commercial reasons, as already mentioned) and does not present any example sentences or phrases.

Among all the dictionaries discussed here, the *CCELD* can be considered as the one that places greater emphasis on the uses of *yes* and *yeah* in conversation, a term that they use in their descriptions. In the entries it provides, they refer to *yes* and *yeah* in both conversation and speech; they also refer to emotive uses of these words (e.g., "[…] *to indicate your involvement* […] "; *"to encourage someone to* […]"). They also mention that the senses of *yes* and *yeah* that they present are "the commonest ones", thus highlighting the fact that these are not the only existing senses and uses of these words.

This brief description of some dictionary definitions of *yeah* and *yes* has been presented at this point in order to illustrate the changes that the use of computerised corpora have introduced to language descriptions. We do not aim, however, at discussing the adequacy and/or comprehensiveness of the descriptions encountered in the samples reproduced above since the present work does not constitute a work in lexicography.

As already mentioned, the novel language descriptions available from dictionaries based on computerised corpora have opened a new range of possibilities for language users in general, influencing, in many cases, both ends of the process of language teaching and learning: on the one hand researchers and teachers and, on the other hand, the learners of the target language.

Two of the most important changes introduced by these works are the emphasis on phraseology and the re-thinking of the division between lexis and grammar. The insights obtained from the theoretical background, from the development of such works and/or from their contents shall be of interest to those working in the fields of applied linguistics and language teaching. Therefore, Chapter 4 will deal with some very important concepts in phraseology, namely collocation, lexical bundles and lexical phrases, since the present study is concerned with the phraseologies that co-occur in *yeah* and *yes* responses.

4- PHRASEOLOGICAL COMPETENCE AND THE TYPOLOGY OF MULTI-WORD ITEMS

4.1- Predictability in discourse and phraseological competence

Studies into collocation, taken in the broad sense of recurrent word combinations (*cf*. AIJMER, 1996, p. 6) or, as we shall refer to it, phraseology (*cf*. COWIE, 1998),' have proved that phrases are an essential part of the workings of the English language. Phraseology, according to Cowie, (*ibid*., p. 1) "has now become the major field of pure and applied research for Western linguists […]", both theoretical and applied, and this is reflected in several international conferences and large-scale research projects.

Much of our knowledge about the use of phraseology can be attributed to some basic concepts related to the organisation of discourse. Spontaneous conversation, according to Stubbs (1983, p. 19), "may look chaotic when closely transcribed [but] is, in fact, highly ordered." The author argues that conversation is polysystemic. In other words, the coherence of spontaneous conversation "depends on several quite different types of mechanisms, such as repetition of words and phrases, structural markers, fine synchronization in time, and an underlying hierarchic structure relating sequences of discourse acts." (STUBBS, *ibid*.)

Stubbs cites Firth, who claimed that

> The moment a conversation is started, whatever is said is a determining condition for what, in any reasonable expectation, may follow. What you say raises the threshold against most of the language of your companion, and leaves only a limited opening for a certain likely range of responses. (FIRTH, 1953 *apud* STUBBS, 1983, p. 84.)

Therefore, the concepts of structure, well-formedness and predictability need to be defined in relation to the notion that the meaning of an utterance depends on its place in a discourse sequence. These concepts can be defined as follows.

> […] the concept of *structure* [is] regarded as constraints on linear sequence. This is closely related to the concept of *well-formedness*: the possibility of distinguishing coherent and incoherent sequences in discourse. It is also closely related to the concept of *predictability*: conversationalists can predict what other speakers are likely to say, because there are constraints on linear sequence. The concept of *structure* is not definable independently of the concept of *system*: different systems of choice are available at different places in structure. Finally, all of these concepts

> are applicable to language only under the assumption that the data are considerably *idealized*. (STUBBS, 1983, p.85, 86.)

Furthermore, Stubbs the concept of well-formedness is described by Stubbs (*ibid.*, p.93) as closely related to the concept of a syntagmatic chain, "in which one item sets up predictions that other items will or will not occur." According to Stubbs (*ibid.*, p. 93, 94),

> It is easy to demonstrate that discourse sequences set up predictions in terms of propositional content.[…] Predictability may be the single most important feature of human communication, precisely since it is central not only to all levels of language, but also central to memory and to thinking in general.

Therefore, the concept of syntagmatic chain and the notion of predictability in discourse are associated by Stubbs to all levels of language, and these features can be considered as fundamental to instances of phrases (or collocation) in language use. In this respect, Kjellmer (*apud* AIJMER and ALTENBERG 1991), notes that, since the mental lexicon of any native speaker contains single-word units as well as phrasal units or collocations, "mastery of both types is an essential part of the linguistic equipment of the speaker or writer and enables him to move swiftly and with little effort through his exposition from one prefabricated structure to the next" (KJELLMER, *apud* AIJMER and ALTENBERG 1991, p. 125).

The predictable nature of the constituents of phrases is described by Kjellmer as a "decisive characteristic," since "the presence of one of them will predict the presence of the other(s)" (*ibid.*). However, the author (who uses the term 'collocation') notes that predictability is not an unconditional aspect of collocation although it is an important element. In other words, collocations may "range from well-established and integrated phrases to doubtfully cohesive sequences of words" (*ibid.*). According to Kjellmer, this implies a continuum in the scale of predictability. At one end, there would be little or no flexibility for the language user to modify a given form; at the other end, the form allows for more creativity or productivity, which, however, would be "all the time conditioned by the selectional restrictions that are defining factors of collocations. (KJELLMER, *apud* AIJMER and ALTENBERG 1991, p. 125.

The arguments presented by Kjellmer for a phraseological (or collocational) perspective of language in use, though having the collocation of words at their root, demonstrate that speakers "move […] from one prefabricated structure to the next" (*ibid.*), thus showing that collocation permeates words, phrases and sequences of words of various types.

In the previous sections we have tried to demonstrate the ways in which phraseological (or collocational) competence is an important feature of language use. The learner whose phraseological competence is still not sufficient is described by Kjellmer (1991) as having as his or her building material "individual bricks rather than prefabricated sections" (*cf*. quotation above). As a consequence of that, we could argue that a new approach to the teaching and learning of EFL would represent one of the possible solutions to learners' problems. According to Kjellmer,

> Lexical items should not be taught and learnt in isolation but only in their proper contexts. This means shifting the emphasis from the individual words to the collocations in which they normally occur. […] Vocabulary learning, from a very elementary level to and upwards, should focus on how the words of the target language are actually used. 'Lexical phrases are in fact basic to language performance' insist Nattinger and DeCarrico (1989: 119). […] This may be particularly important even at fairly elementary levels of vocabulary learning, when the learner is still chiefly preoccupied with high-frequency words. […] It is only when the student has acquired a good command of a very considerable number of collocations that the creative element can be relied on to produce phrases that are acceptable and natural to the native speaker. (KJELLMER, *ibid*., p. 125.)

However, new approaches to the teaching of collocation should not "overstress phraseological knowledge", as research into the practical applications are still at the beginning. Granger (1998, p. 157) notes that, for example, studies into the part played by routine patterns in the development of syntactic structures, in both first- and second-language acquisition, have produced results which are very inconclusive. They seem to indicate that the strategies of using routines and creative constructions develop independently of each other and this view is supported by neurolinguistic evidence, that is, automatic speech has been proved to be neurologically different from creative language. A study by Peters (1977 *apud* GRANGER, *ibid*.) has demonstrated that children use two learning strategies: "analytic" (*i.e.*, from the parts to the whole) and "gestalt" (i.e., from the whole to the parts); however, the "domination of one strategy or the other will depend on individual personality and context of use" (GRANGER, *ibid*.). Yorio's (1989, p. 69 *apud* GRANGER, *ibid*., p.158.) research into adult learners of a second language suggested that these "do not appear to make extensive early use of prefabricated, formulaic language, and when they do, they do not appear to be able to use it to further their grammatical development." Granger (1998) investigated learner's use of prefabs in writing and verified that "learners' phraseological skills are severely limited: they use too few native-like prefabs and too many foreign-sounding ones" (GRANGER, *ibid.*, p.158).

These examples of the very few studies that have been developed into the acquisition of collocation and the development of phraseological competence demonstrate, however, that

the teaching of collocation needs to play a greater role in the teaching of foreign languages than it has in the past. The theoretical background and language studies which use corpora have continuously produced new evidence and insights, whose applications still need further development. Therefore, in any discussion of these issues, it is important to identify a working typology of multi-word items, since the field of phraseology has presented various terms and definitions. This topic will be addressed below.

4.2- A review of the current terminology and typology of multi-word items

In this chapter, we will present a review of the current and most frequently used terminology available for multi-word items. The fields of "lexicology" and "idiomatology" (*cf*. MOON, *apud* SCHMITT and McCARTHY, 1997, p.43) have produced a confusing variety of terms for multi-word items since the various authors have proposed different terms which, sometimes, overlap. In Moon's words (*ibid*., p. 43), there has been "an unruly collection of names for [multi-word items], with confusing results."

Therefore, we shall start this discussion by presenting Moon's (*ibid*.) classification of multi-word items, which, in spite of being incomplete, includes most types of these items "which are relatively well-used or understood in the Anglo and Anglo-American traditions, in preference to more specialist terminology" (*ibid.*). The author also notes that "there is no generally agreed set of terms, definitions and categories in use' (*ibid.*). In this respect, a further problem arises, *i.e*., Moon does not include "lexical phrases" in her typology, although she refers to them in the definition of "prefabs" (*cf*. Moon's own observations in the descriptions of "prefabs" below). However, for most of the present work, we shall refer to Nattinger and DeCarrico's definition of "lexical phrases" (*cf*. Chapter 4.3). The second problem in Moon's typology is the fact that it does not include the more recent "lexical bundles" (*cf*. BIBER *et al*, 1999), which shall be addressed in Chapter 4.4.

A multi-word item is defined by Moon (*ibid*.) as "a vocabulary item which consists of a sequence of two or more words (a word being simply an orthographic unit)" which "semantically and/or syntactically forms a meaningful and inseparable unit." An important feature of multi-word items is that they result from fossilisation and word formation, rather than "the operation of grammatical rules". Therefore, words which have emerged from inflectional processes, such as comparative adjectives and verbs in the passive, cannot be included in this category (*ibid.*).

Moon proposes three important criteria which distinguish holistic multi-word items from other kinds of strings. They are institutionalisation, fixedness, and non-compositionality[31]:

> **Institutionalisation** is the degree to which a multi-word item is conventionalised in the language: does it recur? Is it regularly considered by a language community as being a unit? Pawley (1986) discusses the process and fact of institutionalisation or, in his terms, 'lexicalisation'.
> **Fixedness** is the degree to which a multi-word item is frozen as a sequence of words. Does it inflect? Do its component words inflect in predictable or regular ways? For example, *they rocked the boat* and *not they rock the boated* or *they rocked the boats*. Similarly, does the item vary in any way, perhaps in its component lexis or word order? For example, *another kettle of fish* and *a different kettle of fish* are alternative forms, but *on the other hand* is not varied to *on another hand* or *on a different hand*.
> **Non-compositionality** is the degree to which a multi-word item cannot be interpreted on a word-by-word basis, but has a specialised unitary meaning. This is typically associated with semantic non-compositionality: for example when someone *kicks the bucket* (i.e., 'dies') they are not actually doing anything to a receptacle with their foot, and *cat's eyes* (luminous glass beads set into the road surface to guide drivers) in British English, are not, in any degree biological. However, non-compositionality can also relate to grammar or pragmatic function. For example, *of course* is non-compositional because it is ungrammatical, and the imperative valediction *Take care!* can be said to be non-compositional because of its extralinguistic situational function or 'pragmatic specialisation' (*ibid*., p. 44).

The author notes that these criteria operate together and that in spoken English they are accompanied by a phonological criterion which turns multi-word items into "single tone units". The author also emphasises the fact that all these criteria vary in nature and degree in each multi-word unit (*ibid*.). These units were classified by Moon into five categories: a) compounds, which may "differ from single words only by being written as two or more orthographic words. They cannot properly be separated out altogether". Some examples are *carpark* and *dining-chair*; b) phrasal verbs, which "are combinations of verbs and adverbial or prepositional particles". Examples include *give up* and *call off*; c) idioms, which "refer to multi-word items which are not the sum of their parts: they have holistic meanings which cannot be retrieved from the individual meanings of the component words". They can be exemplified by *spill the beans* and *kick the bucket*; d) fixed phrases include a number of multi-word items which "fall outside the previous categories". Examples include *of course, at least*, *in fact*, *by far, you know, good morning*, and others; e) prefabs or prefabricated routines, which are "preconstructed phrases, phraseological chunks, stereotyped collocations, or semi-fixed strings which are tied to discoursal situations and which form structuring devices".

[31] Despite the long citation that follows, we have opted for reproducing the complete taxonomy proposed by Moon, in order to avoid any innacuracies of the author's definitions in a paraphrase.

Examples: *the thing/fact/point is, that reminds me, I'm a great believer in …* and others. They are often used as specific types of conventions, but may vary and are not "completely frozen". Moon claims that "their non-compositionality stems from their discoursal uses, since their surface meanings can be readily decoded" (*ibid.*, p. 44-47).

Moon draws our attention to the fact that 'there are inevitably overlaps between the categories' (*ibid.*, p. 47). She exemplifies this through the question *what are you driving at?*, in which the forms of either a phrasal verb or a prefab might be identified. Furthermore, the study of multi-word items through the use of corpora, has allowed for the verification, for instance, that idioms "show remarkable degrees of variation" (MOON, *ibid.*, p. 52). In a study of the stability of the forms of multi-word items, the author discovered that forty per cent of the idioms she investigated regularly varied and were unstable in form ("this figure does not include deliberate, jocular, or ad hoc exploitation of idioms as in puns"; *ibid.*).

This suggests, in the author's opinion, that "in any new model of idiom", it might be better to have a notion of "preference of form" or "preferred lexical realisation" rather than "fixedness of form". Consequently, Moon argues that it is preferable to adopt the notion that there is a complex relationship between deep semantics and surface lexis, "rather than it all being a simple case of individual anomalous strings with non-compositional meanings" (*ibid.*, p. 53).

Another important aspect that the author highlights is that, when investigating the use of multi-word items in discourse, we should bear in mind the notion of genre preferences: "[…] by looking at the densities of different kinds of multi-word item in particular text types, it can be seen that there are often strong genre preferences" (*ibid.*, p. 53-54). For example, McCarthy and Carter (*apud* MOON, *ibid.*) identified the frequent use of idioms in horoscopes in journalism. In addition to that, Moon (*ibid.*) identified the frequent use of compounds in technical writing; fixed phrases and idioms (among others) in a report on a soccer match; and compounds, phrasal verbs and prefabs in a screenplay (which attempts to replicate natural speech patterns).

These findings also demonstrate, among other things, that "multi-word items have important roles with respect to the structure of text" (*cf.* MOON, *ibid.*, p. 56). For instance, compounds "typically denote and have high information content", since they are often "technical terms or have specific reference"; fixed phrases and prefabs have been found to "organise and provide the framework for an utterance or the argument of a text; or they are situationally bound" (in cases of ritualistic formulae of greeting, thanking, and others); as to idioms, these "typically evaluate and connote, and are shorthand, rhetorically powerful ways

of conveying judgements". They often play specific discourse roles, such as prefaces or summarisers (*cf*. MOON, *ibid.*).

As already mentioned, Moon's typology of multi-word items and lexical phrases (which she identifies as encompassing prefabs and other multi-word items) is not equivalent to the one provided by Nattinger and DeCarrico (1992); her typology does not include, either, a description of "lexical bundles", which first appeared in BIBER *et al* (1999). Therefore, we shall describe these two types of multi-word items in the next sections.

4.3- Lexical phrases: definition of the term, and their role as part of the grammatical and pragmatic competences

The term "lexical phrase" is not used consensually in the specialised literature. For the purposes of the present study, the definition we shall adopt is the one introduced by Nattinger and DeCarrico (1992). According to the authors, lexical phrases are

> lexico-grammatical units [...], 'chunks' of language of varying length, phrases like *as it were, on the one hand, as X would have us believe,* and so on. As such they are multi-word lexical phenomena that exist somewhere between the traditional poles of lexicon and syntax, conventionalized form-function composites that occur more frequently and have more idiomatically determined meaning than language that is put together each time. These phrases include short, relatively fixed phrases [...] or longer phrases or clauses [...]. Each is associated to a particular discourse function, such as expressing time [...] or relationships among ideas [...]. (NATTINGER and DeCARRICO, 1992, p.1.)

In the authors' view, both grammatical and pragmatic competences are involved in the use of lexical phrases since these are forms of the lexicon which are also a part of grammatical competence. However, the authors argue that "the relationship these forms have to particular functions in context is a matter of pragmatic competence" (*ibid*., p. 7). The authors finally argue that although grammatical competence encompasses the knowledge of the lexical forms and their internal syntax, pragmatic competence accounts for "the speaker's ability to continue to access these forms as pre-assembled chunks, ready for a given functional use in an appropriate context" (NATTINGER and De CARRICO, *ibid*., p. 13).

In their functional classification of lexical phrases (henceforth referred to as LPs) these are divided into three major groups, namely: (1) *social interactions*, (2) *necessary topics*, and (3) *discourse devices*. For the purposes of the present study, the numbers (1), (2) and (3) and the sub-categories (1a) and (1b), used by Nattinger and DeCarrico, shall be retained.

Therefore, all reference to these throughout this work will be made through the use of the same numbers.

Social interactions, or (1), are lexical phrases that are markers describing social relations. Category (2), *necessary topics*, are lexical phrases which characterize topics that are necessary in daily conversations. Category (3), *discourse devices*, are types of lexical phrases that connect the meaning and structure of the discourse.

Some examples of social interaction markers (1), categories of conversational maintenance, include "summoning" (*Hey/hi/hello [NAME]; [NAME] How are you?*), "responding to summons" (*Hello/hi [NAME]; [I'm] Fine, thanks, [and you?]*); "shifting turns" ([*Well,] So OK*); "clarifying (1) audience" (*Excuse me?*); "clarifying (2) speaker" (*What I mean is X.*); and others. Social interaction markers (2), categories of conversational purpose, include "asserting" (*it is [a fact that/the case that] X; I believe [that] X*); "answering" (*Yes, [there /it/they is/are not*] (*X*]); complying (*of course; sure [thing*]); refusing (*of course not; no way*); and various others. Necessary topics include, among others, "autobiography" (*I'm from X; My name is X*); "quantity" (*How much/big is X?; lots of X*); likes (*I like/enjoy X [a lot]; I don't like/enjoy X [at all]; I'd like to X*); and others. Discourse devices include "logical connectors" (*because [of] X; as a result [of X]*); temporal connectors (*and then; after X then; the next is Y*); fluency devices (*and so on; you know; if you see what I mean*); exemplifiers (*for example; it's like X*); evaluators (*at all; at least; I guess*); and various others.

The complete list of categories for the classification of lexical phrases and their respective examples shall be presented in Chapter 5 because they will be used in our analyses of the utterances in the corpus.

4.4- Lexical bundles: definition of the term, and their role in discourse

Lexical bundles are described by Biber *et al* (1999, p. 989) as "extended collocations" since they are word forms which often co-occur in longer sequences, "bundles of words that show a statistical tendency to co-occur." The authors stress the importance of emphasizing the difference between idioms and lexical bundles. The former are relatively invariable expressions with a meaning not derivable from the parts and "not necessarily common expressions at all", whereas lexical bundles are sequences of words that most commonly co-

occur in a register. They are not fixed expressions, usually, and it is not possible to substitute a single word for the sequence (unlike idioms, which, in many cases, can be substituted by a single word with a similar meaning). Most lexical bundles are not structurally complete at all and they frequently "bridge two structural units; in many cases the last word of the bundle is the first element of the second structure (such as in *the fact that* and *I don't know what*; *ibid.*, p. 994-995). They are also described as systematic patterns of use that can only be identified through large-scale corpus studies.

Some examples of lexical bundles used in conversation include sequences such as *do you want me to, I said to him, I don't know what, going to be a*, and various others. Examples encountered in academic prose include *in the case of the, there was no significant, it should be noted that*, and others.

In the scope of the work by Biber *et al* (*ibid.*), the authors decided to investigate lexical bundles of three or more words, in order to make their investigation "more manageable", and argued that "shorter bundles are often incorporated into more than one longer lexical bundle". For example, the three-word lexical bundle *I don't think* is used in many four-word bundles, such as *but I don't think*, *well I don't think*, *well I don't think*, *I don't think so*, and *I don't think I.* Another important characteristic of lexical bundles is the fact that they "must recur frequently in order to be considered a lexical bundle (*ibid.*, p. 990). Furthermore, they differ from other combinations of words that are often repeated within the span of a single discourse in that they are widely used across texts and do not show any variation in form (a lexical bundle represents a specific recurrent word combination).

The authors argue that their study into lexical bundles has revealed that "grammar is not just a study of abstract classes and structures, but of particular words and their particular functions within those classes and functions" (*ibid.*, p. 990). This notion, according to them, is also important for the learner of EFL since "producing natural, idiomatic English is not just a matter of using well-formed sentences, but of using well-tried lexical expressions in appropriate places" (*ibid.*).

The discussions that have been presented in this chapter have focused on aspects of phraseology related to phraseological competence, the acquisition of phraseological competence by learners, the typology used in the field, and a special focus on two types of phrases, *i.e.*, lexical phrases and lexical bundles. These considerations shall be of use for the upcoming chapters, since we will present the methodology (*cf.* Chapter 5) and the analyses (*cf.* Chapter 6) of the phraseological units encountered in the utterances in the corpus under investigation.

5- METHODOLOGY

5.1- Describing the corpus

In the present study I have used data collected from one of the largest corpora of spoken English currently available, namely the Bank of English corpus[32], which is developed by the COBUILD Project (*cf*. Chapter 1 of the present study) of the University of Birmingham. The Bank of English corpus contains instances of both spoken and written English. It encompasses over two hundred million words and is constantly growing (*cf*. BIBER, CONRAD and REPPEN, 1998). The spoken component of the corpus consists of over sixty million words and contains transcripts from spontaneous, informal conversation from all parts of Britain, and radio broadcasts from American and British radio (*cf*. HUNSTON and FRANCIS, 2000, p. 282).

The data which has been analysed in the present study consists of all the utterances containing the words *yeah* and *yes* in the spoken corpus. The *yeah* and *yes* nodes of the corpus (*i.e*., the word in the centre of each line; see the next paragraph a description of the nodes) were selected and all the utterances in them were exhaustively analysed. It is estimated that the *yeah* and *yes* corpora contain 10, 500 words each, thus totalling 21, 000 words approximately.

A *yes* response *(YYR)* has been initially defined in the present study as any response containing the words *yeah* or *yes* at initial, medial or final position. According to Sinclair (1991, p. 143) the word *yes* appears among the top 113 forms in the "COBUILD frequency count" at the 81st position, which represents a very high frequency. The words *yeah* and *yes* also appear in the CANCODE corpus (developed by the University of Nottingham; *cf*. Chapter 3) among the 50 most frequent words from 333,000 words of spoken data (*yeah* appears in the eighth position and *yes* in the twenty-fourth position).

The database of YYRs extracted from the Bank of English corpus was obtained in the form of concordance sheets. A concordance is "a collection of the occurrences of a word-

[32] As already mentioned in Chapter 1, a corpus is "a collection of naturally-occurring language text, chosen to characterize a state or variety of a language." (SINCLAIR, 1991, p. 171).

form, each in its own textual environment" (SINCLAIR, 1991, p. 171). They are organised by a concordancer, which is "[…] a program that searches a corpus for a selected word or phrase and presents every instance of that word or phrase in the centre of the computer screen, with the words that come before and after it to the left and right.[…]", according to Hunston (2002, p. 39).

The concordance sheets were obtained in the KWIC format (Key Word in Context), in which "the word-form under examination appears in the centre of each line, with extra space on either side of it [...]. The full KWIC format prints a whole line of text, with the word under examination in the middle." (*ibid*.) The word in the centre of each line is called the *node, i.e.*, "the central lexeme surrounded by a fixed amount of language in concordance sheets or screens from language corpora", as defined by Crystal (1995, p. 160). This "fixed amount of language" which surrounds a node is called the *span*, defined by Sinclair (1991, p. 175) as "[…] the measurement, in words, of the co-text of a word selected for study". The co-text of a selected word or phrase thus "[…] consists of the other words on either side of it" (*ibid*., p. 172), being "[…] the linguistic environment of any expression under scrutiny" (*ibid*., 171) and it differs from the span in that the latter is a measured piece of verbal context (*cf*. Appendix 9.1 for sample concordance pages for the nodes *yeah* and *yes*).

Sinclair (1991, p. 42) argues that "The quality of evidence about the language which can be provided by concordances is quite superior to any other method; automatic concordancing of texts has been an established facility for many years now […]." He highlights that

> […] the ability to be exhaustive is one of the principal features of a concordance, because it can claim to present all the available information, and is clearly superior to a list of selective citations where there are no strict rules about selection. […] (SINCLAIR, *ibid*., p. 43.)

The concordances below are samples that have been extracted from the *Bank of English* corpus for the nodes *yes* and *yeah*. (The responses are preceded by the symbols < >).

- if they don't want our Queen that's up to them. < > Yes not for that one man though surely
- It's up near the Hagley Road. < > Oh of course it is. Yes it's not as convenient as the Vale site.
- tried a management before at Tranmere hasn't he? < > Er yeah yeah he had a player manager's job there # But

- we'll see what our listeners think but thanks for < > Yeah probably others have got different opinions. The woman that

The initial analyses of the nodes *yeah* and *yes* consisted of the identification of the illocutionary values of the initiation moves, the *yeah* and *yes* responses they originated, and their continuations by the same speaker. As already mentioned in Chapter 1, the illocutionary value of an utterance is the communicative force of an utterance or the "type of function of language" (AUSTIN, 1962, p.100). The term "continuation" has been adopted by some authors (*cf*. STUBBS 1983; AIJMER 1996) and refers to utterances produced by speaker 2 (*i.e.*, the speaker holding the second or responding turn) after uttering an initial *yeah* or *yes*. Continuations are, thus, extensions of a phrase used in responding, and may consist of expansions of the topic or new topics introduced by the same speaker. Continuations may occur in inverted order in *yeah* and *yes* responses, since *yeah* and *yes* may occur in medial or final position, too.

The responses contained the words *yeah* or *yes* alone, at initial, medial or final position. The instances of *yeah* or *yes* responses which did not contain any initiation moves[33] and/or whose illocutionary values could not be identified were considered as invalid. All valid responses were numbered in the corpus and then were rewritten in dedicated forms (*cf*, Appendix 9.1). Each pair of initiation move (IM) and response was written in one form, and their illocutionary values were analysed and annotated.

As mentioned in the preceding paragraphs, the classifications of the illocutionary values of the initiation moves and the *yeah* and *yes* responses in the corpus were analysed and revised at several stages. After the final classification, further analyses were made about the types of initiation moves (*i.e.*, statements, questions or commands) and responses (*yeah* or *yes* only, the existence of continuations realised through phrases and/or combinations of phrases, and their patterns, and the identification of the most frequent illocutionary value of responses). For these purposes, other forms were used in the classifications (*cf*. Appendices 9.2, 9.3, 9.4 and 9.5). The identification of the most frequent *yeah* and *yes* responses aimed at identifying, also, those which contained continuations, since one of our aims is to identify the phraseology of the most frequent responses (*cf*. the discussions of patterns and phrases in Chapters 3 and 4).

[33] Except for a few instances of responses with no initiation moves, whose illocutionary values could be inferred. These are addressed in chapter 6.

It should be noted here that since the classifications of the illocutionary values included those of the initiation moves and the complete responses as seen in the corpus (*i.e*, including the words *yeah* and *yes*, the preceding and subsequent utterances produced by the same speaker within the same turn), the overall count had a multiplying effect which led to numerous revisions (in fact, this total number has not been identified since it is not one of the aims of the present research).

As Crystal (1995, p. 287) notes,

> *Yes* and *no* are among the most commonly used words in the language, but they are often insufficient to capture the various degrees of affirmation or disclination which we may wish to communicate by way of a response. (...) our social survival requires linguistic competence in a much more extensive repertoire of responses.

The identification of the illocutionary values of *yeah* and *yes* responses in the corpus and the analyses of their discursive, interactive and pragmatic properties were made possible due to the fact that these responses can be read in the corpus in conjunction with their initiation moves and any utterances prior or subsequent to the words *yeah* or *yes* (uttered by the same speaker, where applicable). Thus, the tokens which have been analysed do not consist of isolated sentences, but rather, of moves in actual conversation, which enables the analyst to investigate their pragmalinguistic properties and the organisation of the discourse in relation to the internal structure of the moves. They do not consist, either, of sets of complete exchanges, since the corpus consists of concordance lines.

Stubbs (1983, p. 75) argues that a solution to the problem of identifying illocutionary force resides in discourse-based principles. According to him

> The identification of speech acts is not normally regarded as an empirical and testable matter. […] However, illocutionary force can be tested, and one basic procedure of testing and classification is to study the discourse consequences of candidate speech acts. […] I am proposing tests based on possible syntagmatic sequences: discussion of illocutionary force has usually been based entirely on discussions of isolated sentences, not in a discourse sequence […].

In addition to the fact that the identification of the illocutionary values of the initiation moves and responses in the corpus was made possible because both of them are displayed on the same concordance line, it should be noted that the same analytical criteria was used by the team of lexicographers of the COBUILD Project, who also used the Bank of English corpus. According to Moon (1987, *apud* SINCLAIR, 1987, p.87),

> The role of context is a crucial to the lexicography done at Cobuild: context disambiguates. In continuous discourse, whether written or spoken, true ambiguity occurs rarely, excepts where a writer or speaker deliberately wants to be ambiguous. […] In terms of Cobuild lexicography, context is realised by concordances. For most words, the 50 or so characters that appear on either side of the keyword are sufficient for disambiguation. […] The context provided by the concordance line

> gives clear signals of meaning in most cases, in particular through syntax and collocation, and an interplay of these permits disambiguation.

Another important aspect related to the identification of meaning is the fact that, when interpreting utterances in the limited context of concordance lines, some pragmatic decisions have to be made, in order to account for the most probable meaning in a given context. A parallel could be made at this point with what Moon (*ibid*., p. 90) describes about the work of lexicographers who used the Bank of English corpus. She observes that "pragmatic decisions have to be made in editing: a different dictionary would have made different use of the information in the database."

However, since the span in the concordance sheets consisted of approximately ten words before and ten words after the node, which is considered a large span (*cf*. CRYSTAL 1995, p.161), it was possible to identify the most probable illocutionary value of each utterance, despite the fact that we did not have any other additional information about context, nor any prosodic features that might interfere with the decisions that were taken. In addition to that, we have to bear in mind the fact that these decisions are feasible as they are based on our communicative competence, to use Hymes ´s (1972) concept.

Since affirmative responses in English frequently include the use of the words *yeah* and *yes*, I developed a pilot study in which the aim was to identify other types of affirmative responses in natural spoken discourse**.** In that study (*cf*. SILVEIRA PEDRO, 1999), I collected data from dialogues used in films and interviews and all the affirmative responses were transcribed. It could be noted that most instances of affirmative responses, including those in which *yes* or *yeah* were not used, could fit into the general structure of a YYR and their continuations (*cf*. Appendix 9.7 for the complete list of the responses encountered in the pilot study). Therefore, since YYRs are also the types of affirmative responses that are typically exploited in coursebooks (mainly under the form of short answers), I decided to investigate their occurrences in the Bank of English corpus for the purposes of the present study.

In addition to that, the attempts that were made at obtaining concordance pages for lexical phrases and clauses that perform affirmative responses (of the types I encountered in the pilot study) proved impracticable, for various reasons. Firstly, when attempts were made for extracting concordances whose nodes (see this chapter for an explanation of 'nodes') were such phrases, the Bank of English corpus did not recognise nor supply them as nodes, since those phrases consisted of sequences of words; secondly, when a few instances of shorter lexical phrases were obtained as nodes, these appeared in the concordance pages surrounded

by co-texts of utterances whose illocutionary values could not be discerned, and, thus, the illocutionary values of the phrases could not be identified either; thirdly, in all of the previous situations, the identification of the type of move (*i.e.*, an initiation move or a response) could not be identified either. All these factors demonstrated that the use of the *yes* and *yeah* nodes were the most adequate ones for the purposes of the present research, since they allowed for the identification of both the IM and the response, their illocutionary values, and their pragmalinguistic features, as well.

After the classification of all the utterances in the corpus was completed and quantified, and the incidence of the illocutionary values of the IMs and YYRs was identified, the analyses proceeded to investigate the collocational patterns of *yeah* and *yes*, with an emphasis on the phraseology of *confirming* responses, which had the highest frequency among the YYRs with continuations. The investigations into the phraseology of *confirmings* included the analysis of the sequences of lexical phrases, the most frequent verbs and verb tenses, and the collocational patterns of cohesive items (*cf*. one of the forms used in Appendix 9.7).

5.2- The *yeah* and *yes* nodes: two sub-corpora

5.2.1- The *yeah* sub-corpus

The *yeah* sub-corpus (*i.e.*, the concordance pages for the node *yeah*), (henceforth referred to as the *yeah* corpus or YhC) contained a total of 491 *yeah* responses and 484 initiation moves, which totalled 975 utterances. The total number of valid responses was 459 since there were 32 invalid responses (at a frequency of 6.5% of the total number of yeah responses). The valid responses were all those which allowed for the inference of meaning.

Among the valid responses, there were 128 instances of floor returners (*i.e.*, elements which signal to the interlocutor that he/she can proceed in speech, according to LEVINSON, 1983, p. 365). All the instances in which the *yeah* responses were preceded by their initiation moves and whose illocutionary values could be identified were considered valid, except for seven instances of responses whose initiation moves could not be read, but whose contents allowed for the identification of the illocutionary forces of the utterances. Therefore, there was a total of 459 valid responses, which included 128 floor returners, and 331 valid responses when floor returners were excluded.

The valid responses consisted of 253 instances of clause or phrase responses (other than *yeah* alone), 78 *yeah*-only responses and 128 floor returners (as already mentioned in the previous paragraph). Floor returners consisted of instances of *yeah* alone as well.

The figures for the *yeah* corpus are demonstrated in Table 5.1 below.

.

<table>
<tr><th colspan="4">TABLE 5.1 - The YEAH Corpus - distribution of the initiation moves and responses</th></tr>
<tr><td colspan="4">Total number of utterances (initiation moves and responses): 975</td></tr>
<tr><td colspan="4">Total number of initiation moves: 484</td></tr>
<tr><td colspan="4">Total number of yeah responses: 491 (floor returners included)</td></tr>
<tr><td rowspan="2">Invalid yeah responses: 32</td><td colspan="3">Valid yeah-responses: 459</td></tr>
<tr><td>Clause and/or phrase yeah responses: 253</td><td>Yeah-only responses: 78</td><td>Floor returners: 128</td></tr>
</table>

The figures demonstrate that the number of valid *yeah* responses, *i.e.*, 459 instances, represents an incidence of 93.5%. The number of valid *yeah* responses which did not include floor returners totalled 331 instances. These account for 67.4% of the total of 491 responses found in the *yeah* corpus and 72.1% of all valid *yeah* responses.

The clause or phrase *yeah* responses (or other than *yeah* only) totalled 253 instances, thus representing 51.5% of all the *yeah* responses encountered and 55.1% of the valid *yeah* responses. *Yeah*-only responses represented 15.9% of all *yeah* responses, with 78 instances, and 17% of the valid *yeah* responses. Floor returners accounted for 26% of all *yeah* responses, with 128 instances, and 28.3% of the valid *yeah* responses. Therefore the responses that were performed through phrases and/or or clauses were more frequent than the *yeah*-only responses other than floor returners.

5.2.2 – The *yes* sub-corpus

The *yes* sub-corpus (*i.e.*, the concordance pages for the node *yes*; henceforth referred to as the *yes* corpus or YC) contained a total of 502 *yes* responses and 495 initiation moves, which totalled 997 utterances. The number of valid responses was 452 since there were 50 invalid responses (which accounted for 11.6% of the total of *yes* responses.

Among the valid responses, there were 71 instances of floor returners. Thus, all other cases in which the *yes* responses were preceded by their initiation moves and whose illocutionary values could be identified were considered valid, except for seven instances of responses whose initiation moves were not included in the concordances, but whose contents allowed for the identification of the illocutionary forces of the utterances. These shall be addressed in Chapter 6.

The valid responses consisted of 236 instances of clause or phrase *yes* responses (*i.e.*, other than *yes* alone), 145 *yes*-only responses and 71 floor returners (as already mentioned in the previous paragraph). The floor returners consisted of instances of *yes* alone as well.

The figures for the *yes* corpus are presented in Table 5.2 below.

<table>
<tr><th colspan="4">TABLE 5.2 - The YES Corpus - distribution of the initiation moves and responses</th></tr>
<tr><td colspan="4">Total number of utterances (initiation moves and responses): 997</td></tr>
<tr><td colspan="4">Total number of initiation moves: 495</td></tr>
<tr><td colspan="4">Total number of yes responses: 502 (floor returners included)</td></tr>
<tr><td rowspan="2">Invalid yes-responses: 50</td><td colspan="3">Valid yes-responses: 452</td></tr>
<tr><td>Clause and/or phrase yes-responses: 236</td><td>Yes-only responses: 145</td><td>Floor returners: 71</td></tr>
</table>

.

The figures demonstrate that the number of valid *yes* responses totalled 452 instances, which represents an incidence of 90%. The number of valid *yes* responses which did not include floor returners totalled 382 instances. These account for 76.1% of the total of 502 responses found in the *yes* corpus and 84,5% of the total of valid *yes* responses.

The clause or phrase *yes* responses (other than *yes* only) totalled 236 instances, thus representing 47% of all the *yes* responses encountered and 52.2% of the valid *yes* responses.

Yes-only responses represented 29.1% of all *yes* responses, with 146 instances, and 32.3% of the valid *yes* responses. Floor returners accounted for 14.1% of all *yes* responses, with 71 instances, and 15.7% of all valid *yes* responses. Therefore the responses that were performed through phrases and/or or clauses were more frequent than the *yes*-only responses which were not floor returners.

The figures about the token (*cf*. 5.3) utterances encountered the *yes* and *yeah* corpora (in sections 5.1 and 5.2 above) demonstrate that the *yeah* and *yes* responses that are realised by phrases and/or clauses are instances of the language which are of interest for further investigations, due to their pragmatic functions, their interactive nature and the co-occurrence of linguistic items. The figures also demonstrate that the incidences of the different tokens of YYRs encountered are related to the contents of the corpus, which means that the types of utterances that have been encountered will not always occur in the language at the same rates. Furthermore, it should be noted that speakers can also resort to other means of communication, such as gestures and prosodic features when responding. However, our aim resides in the analysis of the spoken language as a means of connecting discourse, and the study of a corpus that is representative of that language will always provide strong evidence of actual instances of language in use.

5.2.3- The frequencies of the *yeah* and *yes* responses in the corpus

The corpora of *yeah* and *yes* responses had similar frequencies of valid responses (*i.e.*, 93.5% and 90% respectively). The types of responses in the two corpora appeared at similar frequencies, too. Clause or phrase *yeah* responses occurred at a frequency of 55.1% among all valid *yeah* responses, whereas clause or phrase *yes* responses appeared in 52.2% of all valid *yes* responses (*cf*. Tables 5.3 and 5.4 below).

Yeah-only responses appeared in 17% of all valid *yeah* responses, whereas *yes*-only responses presented a frequency of 32.3% instances among the valid *yes* responses (*cf*. Tables 5.3 and 5.4 below).

Finally, the frequencies for floor returners had the following figures: they appeared in 28.3% of all valid *yeah* responses and 15.7% of all valid *yes* responses, as demonstrated in Tables 5.3 and 5.4 below.

.

TABLE 5.3 - The *YEAH* Corpus - frequencies of the types of *yeah* responses			
Total number of utterances (initiation moves and responses): 975 Total number of initiation moves: 484 Total number of yeah responses: 491 (floor returners included)			
Invalid *yeah* responses: 6.5%	**Valid *yeah*-responses**: 93. 5%		
	Clause and/or phrase *yeah* responses: 55.1% of the valid *yeah* Rs; (51.5% of all *yeah* Rs)	***Yeah*-only responses**: 17% of the the valid *yeah* Rs; (15.9 % of all *yeah* Rs)	**Floor returners**: 28.3 % of the the valid *yeah* Rs (26 % of all *yeah* Rs)

.

TABLE 5.4 - The *YES* Corpus – frequencies of the *yes* responses			
Total number of utterances (initiation moves and responses): 997 Total number of initiation moves: 495 Total number of *yes* responses: 502			
Invalid *yes*-responses: 10%	**Valid *yes*-responses:** 90%		
	Clause and/or phrase *yes*- responses: *52.2% of the valid yes* Rs; (47% of all *yes* Rs)	***Yes*-only responses:** 32.3% of the valid *yes* Rs; (29.1% of all *yes* Rs)	**Floor returners:** 15.7% of the valid *yes* Rs; (14.1% of all *yes* Rs)

Tables 5.3 and 5.4 above demonstrate that the figures encountered for both corpora indicate that the instances of *yeah* were associated to longer types of responses (*i.e.*, clause or phrase responses) at a similar frequency to that of *yes* responses (55.1% and 52.2%, respectively). These tables also demonstrate that *yes* only responses were more frequent (32.2%) than *yeah* only (17%). However, *yeah* floor returners were more frequent (28.3%) than *yes* floor returners (15.7%).

The data in the descriptions above were instrumental to the analyses that followed them since they presented the overall distribution of the different types of responses (and their initiation moves) which were analysed for the nodes *yeah* and *yes* of the Bank of English

corpus. The analyses of the different types of initiation moves and responses will be presented in Chapter 6.

5.3 –Describing the tokens in the *yeah* and *yes* corpora and identifying pragmatic categories

The descriptions of language that have been developed in the present study deal with actual tokens of language use. Token descriptions of the type implemented in the present study reveal, in Widdowson's (1990, p. 75) words,

> The relative frequency of forms and their habitual co-occurrence in different contexts. [...] A token description might well reveal that some of [the structures presented in a type description] were of rare occurrence, or restricted to a realisation through a limited range of lexical items, almost exclusively confined to certain contexts, or associated with certain meanings. […] token descriptions on a massive scale are now possible with the development of the computer. The evidence they yield does not just quantify the token occurrence of existing category types, for the most part derived from intuition, but also suggests that the types themselves stand in need of revision so that the language as abstractly conceived by the linguist is brought into closer correspondence with the language as actually realized by the user.

Conversely, type descriptions "present a comprehensive array of structures, and give each of them equal descriptive weight; [...] [they] consider language as abstract knowledge" (*ibid*.), while the token descriptions consider it as "actual behaviour" or "as actually realized by the user" (*ibid*.). It would be reasonable to suppose, therefore, that this approach would be in consonance with the pragmatic view of language proposed by Wittgenstein (*cf*. Chapter 2 of the present study).

Widdowson cites Sinclair and his observations about the recent developments in language description, made possible through the use of the computer, and which require "a radical revision of principles of descriptive procedure in general" (*ibid*.).

Within this perspective, the classification of both the IMs and the YYRs consisted of the analyses of the tokens found in the corpus. The YYRs have also been classified in relation to the occurrence or not of lexical phrases. The instances of lexical phrases have, initially, been classified according to the model devised by Nattinger and DeCarrico (1992). In this respect, the present study does not consist of a strict corpus analysis of collocation, since collocation has as its usual "measure of proximity a maximum of four words intervening" (*cf*. SINCLAIR, 1991, p. 170). The present work has gone beyond studies in collocation since it investigates stretches of language which have been initially analysed and classified according

to their illocutionary values[34] and the illocutionary values of the utterances that originated them.

The analyses have been based on the contents of the transcriptions, which also include markers of hesitation and/or pause, laughter, reference to gender and others. This way, some pragmatic features could be detected but not all of them (for instance, there is no further indication of contextual features such as institutional settings, social distance or others in the corpus), since the corpus consists of concordance pages and not of an ethnographic description. The punctuation that was used in the original corpus has been maintained in the examples in the Appendices. The punctuation, in some instances, did not contain interrogation marks or full stops; however, the senses of the utterances could be identified due to information found in the co-texts. As Levinson (1983, p. 53) argues,

> we can compute out of sequences of utterances, taken together with background assumptions about language use, highly detailed inferences about the nature of the assumptions participants are making, and the purposes for which utterances are being used. In order to participate in ordinary language use, one must be able to make such calculations, both in production and interpretation. This ability (...) is based for the most part on quite regular and relatively abstract principles.

A similar process can be found in corpus analysis, as described by Sinclair (1991). His analyses of collocation were data-driven and conducted according to the meanings inferred from the data encountered in concordance lines. In many passages, he made clear that he was providing his personal interpretations of the senses of the collocations under study, as in the following.

> Number 31: I interpret this example as having as direct object *some fifteen twenty kilometres*, rather than being an intransitive with (...).
> Number 22: I interpret this in the sense of 'abandon'. This is not a sense (...). (SINCLAIR, 1991, p. 58).

The author also noted that a relatively small number of instances of a token is enough for the identification of the typical patterns of its realisation in the language, as in the following excerpt, about the word-forms *yield*, *yielded*, *yielding*, *yields* and their respective number of instances in the corpus (51, 25, 20 and 29, respectively)

> this number of instances should be enough to outline the meaning and use of a word [...] (1991: 54); the new choice opened up by the computer is to evaluate actual instances and select the most typical. A complete set of typical instances should exemplify the dominant structural patterns of the language without recourse to abstraction, or indeed to generalization. (*Iibid*., p. 103).

[34] The illocutionary value or force of an utterance is "the communicative value assigned to it as the performance of an illocutionary act (*i.e.*, that part of the speech act which involves doing and not just saying something)," according to Widdowson, 1996, p. 128).

Although a lot can be inferred from the contexts provided by the pairs of IMs and YYRs as well as from the external elements transcribed in the corpus, which are instances of actual language use, the classification of the illocutionary values proposed here can be more safely approached if considered in light of the utterances' idealised meanings.

As Hunston (2002, p. 65-66) argues,

> Concordance lines present information; they do not interpret it. Interpretation requires the insight and intuition of the observer […] human judgement is needed [for the distinction of meanings]. […] the enormous benefit […] is that the human eye can perceive features of language that were hitherto unguessed-at.
> […] looking at the lines themselves suggests […] the most likely interpretation, but it is important to recognise that this is an interpretation of evidence, not 'fact'.

5.4- The functional classification of the lexical phrases in the corpus

As seen in the previous chapter, the functional classification of lexical phrases made by Nattinger and DeCarrico (1992) divides them into three major groups, namely: (1) *social interactions*, (2) *necessary topics*, and (3) *discourse devices*. For the purposes of the present study, the numbers (1), (2) and (3) and the sub-categories (1a) and (1b), used by Nattinger and DeCarrico, shall be mentioned when referring to these categories.

The scheme for analysis of the lexical phrases in the corpus is presented below. We retained Nattinger and DeCarrico's (*ibid*.) scheme and added some new categories and examples. Therefore, although the authors had, for example, one category named "responding", we added new categories which were used in the responses in the corpus, but retained their original category in the scheme for the purpose of demonstrating the original scheme. The asterisked items are lexical phrases and examples which we encountered in the corpus and added to the original scheme (*cf*. Chapter 6 for a discussion of the findings). The categories are accompanied by examples and are presented in alphabetical order.

Categories of lexical phrases:

(1) Social interaction markers consist of

(a) categories of conversational maintenance, and

(b) categories of functional meaning relating to conversational purpose. Some examples are given below.

(a) Conversational maintenance (regularities of conversational interaction that describe how conversations begin, continue, and end).

Summoning:

Excuse/Pardon me; Hey/hi/hello (NAME); Hello, I'm + NAME;
Good morning/afternoon/evening, (how are you?); I didn't catch/get your name
What's up?; (NAME) How are you (doing)?;
Do you live around here?

Responding to summons:

Hello/hi (NAME); Hello, I'm + NAME; How are you (doing)?;
(I'm) Fine, thanks, (and you)?; What's going on/happening?;
*YES (X) / YES?

Nominating a topic:

What's X?; (By the way) Do you know/remember X?;
Have you heard about X?

Clarifying:

(1) audience:

Excuse/pardon me?; What did you mean by X/ when you said X?
*Actually, *In fact,

(2) speaker:

*Actually; How shall I put it?; *In fact;
Let me repeat;
What I mean/I'm trying to say is X;

Checking comprehension:

All right?; (do you) understand (me)?

Shifting a topic:

*I'm glad you mentioned X; Where was I?; Oh that reminds me of X;
*OK, now, X; What I was trying to say;
*To get back to what I was saying; (Say,) By the way;

Shifting turns:

(Well,) So OK; Excuse me/pardon me; Could I say something here?

Closing:

Well, that's about it; I've got to go / run / do X; I must be going;
(It's been) nice talking to you; I mustn't keep you any longer;

Parting:

Goodbye; See you later; (Well) So long (for now)

b) Conversational purpose (types of speech acts; *i.e*. functions that describe the purposes for which conversations take place).

***Accepting**[35]**/ *Acknowledging / *Agreeing:**

Yeah/Yes (X) (that's) (all) right; *Yes, but (I think that) X;

***Advising:**

Don't worry; If I were X, I ____

Asserting:

there is/are/was/were X; *I think/ believe (that) X; in my opinion; it seems (that) X; it sounds X; it/ *that is (a fact/the case that) X; it's said that X; word has it that X

Answering:

Yes, (there/it/they is/are not) (X) *Yes (alone) *(short answers)

***Apologising:**

(I'm) sorry

***Comparing:**

(a/the) More Adj.; ____ as (much) (as); X is/was/were better/worse (than Y); (X) is/was/were like (Y)

Complimenting:

NP + be/look + (intensifier) + Adj;

I + (intensifier) + like/love + NP

Complying:

I'd be glad/happy to; of course;

sure (thing); * Yeah / Yes (X);

***Confirming**:

Yeah / Yes (X)

***Contradicting:**

*But S+V+O; *but erm

Denying:

X did no wrong

[35] The asterisked categories and items have been added to Nattinger and Decarrico's scheme.

*__Endorsing__:

Yeah / Yes (X)

Expressing gratitude:

Thanks (very much/a lot) (for);

I (really) appreciate your thoughtfulness/kindness/doing X

Expressing politeness:

Thanks (very much); (please,) If you don't mind

Expressing sympathy:

I'm (very) sorry about/to hear (about) X; (wow), That's/how terrible/awful;

What a shame/pity/terrible thing

***Giving opinion:**

S+V+O; Yeah / Yes (X)

*__Giving permission__:

Yes

*__Greeting__:

Yes

***Hypothesizing:**

If + noun + verb

***Informing:**

S+V+O; Yeah / Yes (X)

***Justifying:**

Because (what + V + O) / S+V+O

Offering:

Modal + Pro + VP (i.e. May/Can I help (you)?; Would you like X?)

Questioning:

(rising intonation) Do you X?; Is/are there/it/they X?

Refusing:

of course not; no way; I'd rather you X; I'm sorry but (I'm afraid/ I think that) X;

Requesting:

*Do/Have + you + Verb Phrase

* Imperative verb (X);

Modal + Pro + Verb Phrase (*i.e.* Would you (mind) X?; May I X?)

8Yes?

Responding:

(1) **acknowledging:**

*Yeah / Yes; *Go on;

(Simple reinforcers): (and then) What happened (next/then/after that)?

(2) **accepting:**

*(Oh,) yes; (Yeah,) (I know);

(Oh,) I see, no kidding

(3) **endorsing:**

*Yeah / Yes, (that's) (all) right; *I absolutely/certainly/completely agree;

(That's) a (very) good/excellent point;

There you go;

That's great;

*I (do) remember that/it;

* It really/surely X;

*Sure;

*I suppose we could;

*Lovely;

*No problem

(6) **disagreeing*:**

I don't (really/quite) agree (with you/X)

***Suggesting:**

(perhaps) X can/could; *Why don't you X?(#4)

(2) Necessary topics: these lexical phrases mark topics that are necessary in daily conversations, and ones about which Ls are often asked. The following are some examples.

Autobiography:

My name is _____; I'm from ____; I'm (a) ____/(years old)

Language:

Do you speak ____?; how do you say/spell ____?;

I don't speak ____ very well; I speak ____ (a little)

Quantity:

* a bit of ___; How much/big is ____?; lots of ____; (not) a great deal

* plenty of ___;

Time:

a ____ago, since X; at/it's ____o'clock; for a long time/ ___years; *for years

on ____ day; the ____before/after;

When is X; what time X ?;

Location:

Where is ____?; What part of the ____?; across from ____; next to ____;

to the right/left (of ____); how far is ____?; ____blocks (from ____)

Weather:

I'm ____; Is it going to X?; it's (very) ____ (today)!

Likes:

I don't like/enjoy ____ (at all); * I hate; I like/enjoy ____ (a lot); I'd like

to X; ____ is lots of fun; (what) do you like to X?;

Food:

I'd like (to have) ___/to make a reservation (for ____); serve

breakfast/lunch/dinner;

a table for ___

Shopping:

How much is/are ____?; I want to buy/see ___; it (doesn't)fit(s);

(not) too expensive; a (really) good/bad buy/bargain;

_____ cost(s) (me/you/them) ____dollars

(3) Discourse devices are lexical phrases that connect the meaning and structure of the discourse. Some examples are listed below.

Logical connectors:
* and (?); as a result (of X); because (of) X; in spite of X ; nevertheless

Temporal connectors:
and then; *and when; after X then; after X then/ the next is Y;
the day/week/month/year before/after ___; the next is Y;

Spatial connectors:
around here; *here; at/on the corner; over there;

Fluency devices:
as a matter of fact; at any rate; and so on; by and large;
*I think that X; if you see what I mean; ? in fact ?; it seems (to me) that X; so to speak; you know; *you see

Exemplifiers:
for example; in other words; it's like X ; *like that;
to give you an example

Relators:
*and also; *at the same time (#105); *Oh,
*the (other) thing X is Y;
X has (a lot)/doesn't have (much) to do with Y; not only X but also Y;
*? not for X Y though ?;

Qualifiers:
*that's true but; it depends on X; the catch is X; it's only in X that Y;
*Well, yes but

Evaluators:
* a bit; *apparently (127); as far as I know/can tell;
at all; at least;
I don't know; I guess;
*I'm (not) absolutely sure/positive/certain (but) ____ ;
*I'm (not) sure; it could be (that); *of course;
*probably; there's no doubt that;

Summarisers:
*in effect ____; OK (level intonation); *(OK) so ?;
my/the point (here) is (that) X; *the thing is; * the thing (here) is X;
* that's about all there is to it;
to make a long story short

***Particulariser:**
*in that case; * most (of) _____;
*one aspect of _____;*the first/second ____;

The next chapter will present the analyses of the *yeah* and *yes* responses in the corpus. The data used in these analyses excludes the instances of floor returners (*cf*. sections 5.1, 5.2 and 5.3 above).

6- ANALYSIS OF THE PHRASEOLOGY OF THE *YEAH* AND *YES* RESPONSES IN THE CORPUS

6.1- Introduction

In order to investigate if the words *yeah* and *yes* and the responses in which they occur have any phraseologies, we shall begin this chapter with a description of the analyses of the illocutionary values of all the valid *yeah* and *yes* responses and their initiation moves as encountered in the corpus. As already mentioned in the previous chapters, the utterances which were analysed were obtained from the *yeah* and *yes* nodes of the Bank of English corpus. The choice for these nodes allowed for the identification of both the responses and the initiation moves which originated them. It also allowed for the identification of any continuations to the responses as uttered by the same speaker. As already mentioned in the preceding paragraphs, these continuations consist of other utterances that the speaker adds to the words *yeah* and *yes*, which can complement the initial response contained in *yeah* or *yes* or provide additional information, among others, as shall be further discussed in this chapter.

It should be noted that the figures about the two corpora which were presented in Chapter 5 related to all the utterances in them, including the invalid responses and floor returners (*cf*. Tables 5.1, 5.2, 5.3, and 5.4). However, the figures in chapter 6 will deal with the valid IMs and responses, but will exclude floor returners. Thus, the valid initiation moves total 697 instances and the valid *yeah* and *yes* responses total 713 instances, in a total of 1410 utterances.

Section 6.2 below presents the findings for the illocutionary values of the valid initiation moves and responses in the corpus. The analyses for the nodes *yeah* and *yes* are presented, in most cases, as separate items. They are later compared and contrasted.

6.2- The illocutionary values of the *yeah* responses in the corpus and their initiation moves

The *yeah* responses in the corpus were originated by initiation moves which consisted of questions (interrogative form), statements (affirmative form), statements followed by tag questions, and commands (imperative verbs).

The total number of valid *yeah* responses in the corpus is 331 and the number of valid initiation moves is 322, since there were 9 instances of valid *yeah* responses for which no initiation move could be read in the span. The fact that these 9 responses did not contain initiation moves in the corpus did not invalidate them because their illocutionary values could be inferred from the co-text of the responses themselves. However, other instances of absence of the initiation moves invalidated the responses since their illocutionary values could not be identified.

The analyses revealed that 259 (80,4 %) initiation moves consisted of statements (affirmative form sentences). The second most frequent sentence type was questions, with 59 instances (18,3 %). Commands represented the third preference for initiation moves, with 4 instances (1,2 %).

Table 6.1 presents the figures for the illocutionary values of the valid *yeah* responses in the corpus and of their initiation moves.

.

TABLE 6.1- The illocutionary values of the initiation moves and *yeah* responses in the Bank of English corpus (*yeah* node)

.

Responses / Initiation moves	Accepting	Ack-now-ledging	Agree-ing	Com-plying	Con-firm-ing	En-dors-ing	Giving opinion	Informing	Reporting speech	Shift-ing topic	Total no. of IMs
Statements:											
1.Acknowledging					**10**						**10**
2.Apologising		**1**									**1**
3.Complying					**1**						**1**
4.Confirming					**1**						**1**
5.Demonstrating		**2**									**2**
6. Fact (statement of)		**4**			**17**						**21**
7. Giving directions		**1**									**1**
8. Giving opinion		**1**	**67**								**68**
9. Hypothesising			**6**		**2**						**8**
10. Informing	**2**	**67**			**27**	**9**	**1**				**106**
11. Making decision						**2**					**2**
12. Necessity (expression of)		**4**	**4**								**8**
13. Parting		**1**									**1**
14. Predicting			**2**								**2**
15. Seeking agreement			**1**								**1**
16. Seeking confirmation					**19**						**19**
17. Suggesting			**6**								**6**
18. Sympathising		**1**									**1**
(Sub-total:)											**259**
Questions:											
19.Requesting agreement			**3**								**3**
20.Requesting confirmation					**15**						**15**
(Continued)											

.

(Continued from T. 6.1)											
Responses / Initiation moves	Accepting	Ack-now-ledging	Agreeing	Complying	Confirming	Endorsing	Giving opinion	Informing	Reporting speech	Shifting topic	Total no. of IMs
21.Requesting information		2				1 (self)*		28			31
22.Requesting opinion						1 (self)*	8				9
23.Requesting permission				1							1
(Sub-total:)											59
Commands:											
24.Giving instructions				1							1
25.Requesting action				3							3
(Sub-totals:)											
No initiation move						5 (self)			1	1	7
Total number of responses:	2	84	89	5	92	18	9	28	1	1	322 / 331
Total number of utterances:											655

Table 6.1 shows that the IMs that generated *yeah* responses consisted of utterances of 25 different illocutionary values. The IMs which were realised through statements totalled 18 different illocutionary values, namely *acknowledging, apologising, complying, confirming, demonstrating, fact (statement of), giving directions, giving opinion, hypothesising, informing, making decision, necessity (expression of), parting, predicting, seeking agreement, seeking confirmation, suggesting,* and *sympathising* (*cf*. Appendix 9 for examples).

Question IMs totalled five illocutionary values, namely *requesting agreement, requesting confirmation, requesting information, requesting opinion,* and *requesting permission* (*cf*. Appendix 4 for examples).

Commands occurred in the corpus through two different illocutionary values, namely *giving instructions* and *requesting action* (*cf*. Appendix 4 for examples).

The most frequent illocutionary values for IMs in the corpus were distributed as follows: *informing* was the IM that originated most *yeah* responses in the data, with 106 instances; *giving opinion* was the second most frequent one, with 68 instances; and *requesting information* appeared in the third position, with 31 cases, followed by the other illocutionary values which are presented in Table 6.1.

Table 6.1 also demonstrates that the two most frequent IMs that generated *yeah* responses (*informing* and *giving opinion*) were realised through statements. These two IMs account for over 50% of the total of IMs that generated YhRs, whereas the two most frequent question IMs account for 14.3% of all IMs that generated YhRs, and commands account for 1.2% only.

These findings demonstrate that there is a strong prevalence of statements over questions and commands in the generation of YhRs in spoken discourse. Furthermore, the figures indicate that most instances of YhRs in discourse are generated by *informing* IMs. These are followed by *giving opinion* IMs.

Therefore, it could be argued that, in dialogic conversation, speakers resort more frequently to statements (rather than questions) in the exchange or elicitation of information, and that statements frequently generate YhRs in S2's uptake (in addition to various other possible types of responses). Furthermore, the figures demonstrate that when speakers *inform* and *give opinion*, for example, their discourse is jointly constructed by the interlocutor's (affirmative) YhRs very frequently. In other words, these two illocutionary values do not occur in conversation as self-contained items, but are continuously constructed by the two speakers involved. When S2 hears a piece of *information* or S1's *opinion*, he or she usually

resorts to some type of continuation to that information or opinion, rather than, for example, switching immediately to a different topic. This would represent, in general terms, the tendency that speakers have for cooperation in discourse, thus adding their own contributions to the ongoing discourse.

The *yeah* responses in the corpus had 10 different illocutionary values, namely *accepting*, *acknowledging*, *agreeing*, *complying*, *confirming*, *endorsing*, *giving opinion*, *informing*, *reporting speech*, and *shifting topic* (*cf*. Table 6.1 and Appendix 5 for examples). The five most frequent types of YRs were *confirming* (92 instances), *agreeing* (89), *acknowledging* (84), *informing* (28) and *endorsing* (18). All of these values were mostly realised after a statement IM, except for *informing* (28), which was completely realised after a question IM. *Confirming* was generated by statement IMs in 83.7% of cases; *agreeing* in 96.6%; *acknowledging* in 97.6%; and *endorsing* in 61.1%. The illocutionary value of *endorsing* presented a few instances of a speaker endorsing him/herself. Some probable reasons for this are the facts that the speaker was briefly interrupted by another speaker, had produced a series of utterances or wanted to emphasize the subject he/she was talking about.

Differently from the responses uttered after an *informing* IM (which consisted of statements), the *informing* YhRs were all uttered after a question IM, namely *requesting information*. Therefore, it could be argued that interrogative-form *requests for information* very frequently generate YhRs (in the case of responding affirmatively).

The figures thus indicate that *confirming*, *agreeing* and *acknowledging information* are illocutionary values that are largely realized through utterances containing *yeah* in spoken discourse. They present evidence of high frequencies of YhRs which may contradict intuition.

Section 6.3 is about the illocutionary values of the *yes* responses in the corpus and their initiation moves. It should be noted that the two corpora of *yeah* and *yes* responses have been treated separately since they have, from the beginning of the analyses, provided evidence that there are differences between them.

6.3- The illocutionary values of the *yes* responses in the corpus and their initiation moves

The *yes* responses in the corpus were originated by initiation moves which consisted of questions (interrogative form sentences), statements (affirmative form sentences),

statements followed by tag questions, commands (imperative verbs), and *summons*[36] uttered by radio announcers during phone-in radio programmes.

The total number of valid *yes* responses in the corpus is 375 and the number of valid initiation moves is 382. There were 7 instances of valid *yes* responses for which no initiation move could be read in the span. Similarly to what happened in the *yeah* corpus, their illocutionary values could be inferred from the co-text of the responses. All other instances of *yes* responses which did not contain any IMs were, however, invalidated since their illocutionary values could not be identified.

Table 6.2 presents the illocutionary values of the initiation moves and the *yes* responses they generated, and their distribution in the corpus.

[36] Summons consisted solely of the forms *Hello*, *Hello X* and *Yes X*. They were IMs uttered by radio announcers during radio phone-in programmes who signalled to a member of the audience that it was their turn to speak.

.

TABLE 6.2: The illocutionary values of the initiation moves and *yes* responses in the Bank of English corpus (*yes* node)

.

Responses / Initiation moves	Accepting	Acknowledging	Agreeing	Confirming	Contradicting	Endorsing	Giving opinion	Giving permission	Greeting	Informing	Reporting speech	Responding to summons	Requesting confirmation	Shifting topic	Total no. of IMs
Statements:															
1.Acknowledging		**1**		**2**											**3**
2. Accepting offer		**1**													**1**
3. Agreeing				**4**									**1**		**5**
4. Complaining		**1**													**1**
5.Confirming		**2**		**2**											**4**
6. Fact (statement of)		**1**		**5**											**6**
7. Giving directions		**1**													**1**
8. Giving opinion		**2**	**48**												**49**
9.Greeting									**1**						**1**
10. Hypothesising			**3**												**3**
11. Informing		**94**	**2**	**27**	**1**	**22**				**1**			**1**		**149**
12. Offering food	**1**														**1**
13. Offering help	**1**														**1**
14. Parting		**1**													**1**
15. Predicting			**1**												**1**
16. Prompting speech		**1**													**1**
17. Seeking confirmation				**36**											**36**
18. Seeking information										**2**					**2**
19.Suggesting			**1**												**1**
20. Sympathising		**1**													**1**
(Sub-total:)															**268**
Questions:															
21.Requesting confirmation				**29**											**29**
23.Requesting opinion							**15**								**15**
24.Requesting permission								**2**							**2**
(Sub-total:)															**93**
(Continued)															

(Table 6.2 continued)															
Responses / Initiation moves	**Accepting**	**Acknowledging**	**Agreeing**	**Confirming**	**Contradicting**	**Endorsing**	**Giving opinion**	**Giving permission**	**Greet-ing**	**Informing**	**Reporting speech**	**Responding to summons**	**Requesting confirmation**	**Shif-ting topic**	**Total no. of IMs**
25. Summons												**11**			**11**
Commands:															
26. Advising	**3**														**3**
No initiation move						**2** (self)					**2**			**3**	
Total number of responses	**5**	**108**	**55**	**105**	**1**	**24**	**15**	**2**	**1**	**48**	**2**	**11**	**2**	**3**	**375 / 382**
Total number of utterances															**757**

As demonstrated in table 6.2, the IMs in the *yes* corpus consisted of utterances of 26 different illocutionary values. Statement IMs totalled 20 different illocutionary values, namely *acknowledging, accepting offer, agreeing, complaining, confirming, fact (statement of), giving directions, giving opinion, greeting, hypothesising, informing, offering food, offering help, parting, predicting, prompting speech, seeking confirmation, seeking information, suggesting,* and *sympathising* (*cf*. Appendix 6 for examples).

The IMs realised through questions occurred in 4 illocutionary values, namely *requesting confirmation, requesting information, requesting opinion,* and *requesting permission* (*cf*. Appendix 6 for examples).

Commands occurred in the corpus with one illocutionary value only, namely *advising*. Summonings occurred at 11 instances and have been presented as a separate item in Table 6.2 since their form consisted of the formulae *Hello, Hello X* and *Yes hello* (*cf*. Appendix 6 for examples).

The distribution of the most frequent illocutionary values of the IMs in the *yes* corpus is the following: *informing* was the most frequent IM, with 149 instances; *giving opinion* was the second most frequent one, with 49 instances; *requesting information* appeared in the third position, with 47 instances; and *seeking confirmation* occurred 36 times. These are followed other illocutionary values (*cf*. Table 6.2).

The two most frequent IMs that in the *yes* corpus, namely *informing* and *giving opinion,* were realised through statements. These two IMs account for over 52.8 % of the total of IMs that generated YRs, whereas the two most frequent question IMs account for 20.3 % of all IMs that generated YRs; commands account for 0.8 %; and *summonings* represent 2.9%.

These findings demonstrate that, similarly to the IMs in the *yeah* corpus, there is a strong prevalence of statements over questions and commands in the generation of YRs in spoken discourse. Furthermore, the figures indicate that most instances of YRs in discourse are generated by *informing* IMs. These are followed by *giving opinion* IMs. Therefore, as already noted in the case of the *yeah* corpus, it could be argued that speakers resort more frequently to statements (rather than questions) in the exchange or elicitation of information, and that statements frequently generate YRs in S2's uptake (in addition to various other possible types of responses). Furthermore, as already noted about the YhC, the figures demonstrate that affirmative responses containing *yes* (at initial, medial or final position) are used by speakers to respond to someone who provides *information* or *gives opinion*, for

example. This demonstrates, again, that S2 engages, very frequently, in jointly constructing the discourse that stems from their interlocutors through the use of affirmative responses.

The *yes* responses in the corpus had 14 different illocutionary values, namely *accepting, acknowledging, agreeing, confirming, endorsing, giving opinion, informing, reporting speech, shifting topic, contradicting, giving permission, greeting, request confirmation* and *responding to summons* (*cf*. Table 6.2, and Appendix 7 for examples). The four most frequent types of YRs were *acknowledging* (108 instances), *confirming* (105), *agreeing* (55), *informing* (48) and *endorsing* (24). All of these values were mostly realised after a statement IM, except for *informing* (48), which was mostly realised after a question IM (in 93.7 % of cases). *Acknowledging* was generated by statement IMs in 98.1 % of cases; *confirming* in 72.4 %; *agreeing* in 100%; and *endorsing* in 91.7 %. The *informing* YRs were mostly uttered after a question IM, namely *requesting information (*with 43 instances), against two statement IMs *(informing* and *seeking confirmation)*. Therefore, it could be argued that interrogative-form *requests for information* generate YRs more frequently than affirmative forms of elicitation of information.

The figures indicate that *acknowledging information, confirming* and *agreeing* and are illocutionary values that are largely realized through utterances containing *yes* in spoken discourse. They present evidence of high frequencies of YRs which occur in natural spoken discourse.

These initial findings obtained from the two corpora of *yeah* and *yes* responses shall be discussed at the end of this chapter. As already mentioned, the two corpora have been treated separately due to the differences in the evidence they have provided. Some additional considerations about the two corpora shall be presented in section 6.4, in which they will be compared and contrasted in terms of the illocutionary values of the initiation moves and the *yeah* and *yes* responses they contain.

6.4- The illocutionary values of *yeah* and *yes* responses and their initiation moves compared and contrasted

The illocutionary values of the valid *yeah* and *yes* responses and their initiation moves have been grouped in Tables 6.4 and 6.3, respectively. The purpose of presenting tables which gather information about the two different corpora is one of comparing and contrasting the findings.

Table 6.3 (below) demonstrates that there were nine illocutionary values of IMs in the *yeah* corpus which did not appear in the *yes* corpus. Among these IMs, 5 illocutionary values occurred more than once and one particular illocutionary value, namely *expressing necessity*, occurred 8 times, against no occurrences in the *yes* corpus.

The *yes* corpus contained 10 illocutionary values of IMs which did not appear in the YhC. Among these, there were 4 illocutionary values that occurred more than once and 2 particular illocutionary values occurred at higher frequencies, namely *summoning* (with 11 instances) and *agreeing* (with 5 instances).

The figures for the responses in the corpora are contrasted in Table 6.4 below. The *yeah* responses contained one illocutionary value that differed from those of the *yes* responses, namely *complying*, which occurred five times. However, the *yes* responses had 5 differing illocutionary values, 3 of which occurred more than once. Among these, *responding to summons* occurred 11 times, in correspondence with the specific IMs that originated them.

Further discussions about these findings shall be presented at the end of the chapter.

.

TABLE 6.3 – The illocutionary values of the initiation moves in the *yeah* and *yes* corpora

.

***Yeah* corpus**	**Number of**	***Yes* corpus**	**Number of**
Initiation moves	**instances**	**Initiation moves**	**instances**
1. Acknowledging	10	*1.Acknowledging*	3
*2. Apologising**	1	*2. Advising**	3
*3. Complying**	1	*3. Accepting offer**	1
4. Confirming	1	*4. Agreeing**	5
*5. Demonstrating**	2	*5.Complaining**	1
6. Fact (statement of)	21	*6.Confirming*	4
7. Giving directions	1	*7. Fact (statement of)*	6
*8. Giving instructions**	1	*8. Giving directions*	1
9. Giving opinion	68	*9. Giving opinion*	49
10. Hypothesising	8	*10.Greeting**	1
11. Informing	106	*11. Hypothesising*	3
*12. Making decision**	2	*12. Informing*	149
*13. Necessity (expression of)**	8	*13. Offering food**	1
14. Parting	1	*14. Offering help**	1
15. Predicting	2	*15. Parting*	1
*16. Seeking agreement**	1	*16. Predicting*	1
17. Seeking confirmation	19	*17. Prompting speech**	1
18. Suggesting	6	*18. Seeking confirmation*	36
19. Sympathising	1	*19. Seeking information**	2
*20. Requesting action**	3	*20. Suggesting*	1
*21. Requesting agreement**	3	*21. Sympathising*	1
22. Requesting confirmation	15	*22.Requesting confirmation*	29
23. Requesting information	31	*23.Requesting information*	47
24. Requesting opinion	9	*24.Requesting opinion*	15
25. Requesting permission	1	*25.Requesting permission*	2
-		*26. Summoning**	11
Total number of IMs	**322**	**Total number of IMs**	**375**
Total number of differing illocutionary values in the two corpora: 19			

* The asterisked items are the illocutionary values of IMs which differ in the two corpora.

.

TABLE 6.4 - The illocutionary values of the responses in the *yeah* and *yes* corpora (‘*floor returners’* excepted)

.

***Yeah* corpus**	**Number of instances**	***Yes* corpus**	**Number of instances**
Responses		**Responses**	
1. Accepting	2	*1. Accepting*	5
2. Acknowledging	84	*2. Acknowledging*	108
3. Agreeing	89	*3. Agreeing*	55
*4. Complying**	5	*4. Confirming*	105
5. Confirming	92	*5. Contradicting**	1
6. Endorsing	18	*6. Endorsing*	24
7. Giving opinion	9	*7. Giving opinion*	15
8. Informing	28	*8. Giving permission**	2
9. Reporting speech	1	*9. Greeting**	1
10. Shifting topic	1	*10. Informing*	48
-		*11. Reporting speech*	2
-		*12. Requesting confirmation**	2
-		*13. Responding to summons**	11
-		*14. Shifting topic*	3
Total number of responses	**331**	**Total number of responses**	**382**
Total number of differing illocutionary values in the two corpora: 6			

* The asterisked items are the illocutionary values of *yeah* and *yes* responses which differ in the two corpora.

6.5- The general distribution of the responses which contained continuations in the corpus

As already mentioned in Chapter 5, complex responses have been defined in the present study as those which contain clauses and/or phrases, and independent responses are those which consist of *yeah* or *yes* alone (in single or repeated realisations) or independent phrases, *i.e*, short, semi-fixed lexical phrases (containing *yeah* or *yes* as a core) which are used alone in the response. Independent phrases are not followed by any additional utterances by S2.

The responses in the corpus contained, in many instances, clause or phrase continuations to the words *yeah* and *yes.* The figures presented in the previous sections demonstrated the number of valid responses for the two nodes. The valid responses included instances of *yeah* or *yes* only (in single or double realisations) and clauses and/or phrases.

In this section, we shall, firstly, identify the number of instances of responses in the corpus which did and did not contain instances of continuations. The findings are presented in Tables 6.5, 6.6 and 6.7.

Table 6.5 presents the number of instances of *yeah* which were followed by clause or phrase continuations.

.

Table 6.5 – Incidence of continuations uttered after *yeah*

.

Illocutionary value of *yeah* / incidence of instances	**Incidence of instances of continuation**
Accepting: 2	Clause or phrase: 1 None: 1
Acknowledging: 84	Clause or phrase: 54 None: 30
Agreeing: 89	Clause or phrase: 55 None: 34
Complying: 5	Clause or phrase: 5 None: 0
Confirming: 92	Clause or phrase: 64 None: 28
Endorsing: 18	Clause or phrase: 15 None: 3
Giving opinion: 9	Clause or phrase: 8 None: 1
Informing: 28	Clause or phrase: 22 None: 6
Reporting speech: 1	Clause or phrase: 1 None: 1
Shifting topic: 1	Clause or phrase: 1 None:
Total incidence of valid *yeah* responses: 331 (floor returners excepted)	**Total incidence of clause or phrase continuations after *yeah*: 226** **No continuation: 105**

.

.

Table 6.6 – Incidence of continuations uttered after *yes*

.

Illocutionary value of *yes* / incidence of instances	**Incidence of instances of continuation**
Accepting: 5	Clause or phrase: 3 None: 2
Acknowledging: 108	Clause or phrase: 37 None: 71
Agreeing: 55	Clause or phrase: 35 None: 20
Confirming: 105	Clause or phrase: 73 None: 32
Contradicting: 1	Clause or phrase: 1 None: 0
Endorsing: 24	Clause or phrase: 22 None: 2
Giving opinion: 13	Clause or phrase: 13 None: 0
Giving permission: 2	Clause or phrase: 2 None: 0
Greeting: 1	Clause or phrase: 1 None: 0
Informing: 48	Clause or phrase: 34 None: 14
Requesting confirmation: 1	Clause or phrase: 0 None: 1
Reporting speech: 2	Clause or phrase: 2 None: 0
Responding to summons: 10	Clause or phrase: 10 None: 0
Shifting topic: 3	Clause or phrase: 3 None: 0
Total incidence of valid *yes* responses: 381 (floor returners excepted)	**Total incidence of clause or phrase continuations after *yes*: 236** **No continuation: 145**

.

Table 6.7 - The most frequent illocutionary values of the valid *yeah* and *yes* responses which contained continuations

.

Illocutionary value of *yeah*	Incidence of continuations	Illocutionary value of *yes*	Incidence of continuations	Total incidence of continuations
Confirming	64	*Confirming*	73	137
Agreeing	55	*Agreeing*	35	91
Acknowledging	54	*Acknowledging*	37	90
Informing	22	*Informing*	34	56
Endorsing	9	*Endorsing*	21	30
Giving opinion	8	*Giving opinion*	13	21
--	--	*Responding to summons**	10	10
Endorsing self	6	*Endorsing self*	1	7
*Complying**	5	--	--	5

* The asterisked items occurred in one corpus only.

Tables 6.5 and 6.6 demonstrate that the responses in the corpus contained continuations to the words *yeah* (in 226 instances) and *yes* (in 236 instances) at the frequencies of 68.3 % and 61.9 %, respectively. There were no continuations after *yeah* (in 105 instances) and after *yes* (in 145 instances) at the frequencies of 31.7% and 38.1 %, respectively. Therefore, since the figures demonstrate that most responses in the corpus contained continuations, it could be argued that speakers seem to have a preference for providing additional information when they produce *yeah* and *yes* responses in spoken English. Furthermore, if we take into consideration the fact that a certain amount of *yeah* or *yes* alone might occur in conversation due to interruptions by an interlocutor, or due to hesitations, for instance, we can presume that the number of continuations could be even higher.

Table 6.7 presents the most frequent illocutionary values of the responses which contained continuations in both corpora (in decreasing order of frequency). It demonstrates that *confirmings* are the responses that were most frequently followed by continuations, with 137 instances; in the second and third positions appear *agreeing* and *acknowledging*, with 91 and 90 instances respectively. The fourth most frequent one is *informing*, with 56 instances, which is followed by *endorsing* (in the fifth position), with 30 instances, and *giving opinion* (in the sixth position), with 21 instances. The seventh most frequent responses with continuations, *responding to summons*, occurred only in the *yes* corpus. *Endorsing self*

appeared in both corpora, with 7 instances, whereas *complying* appeared only in the YhC, with 5 instances.

The figures demonstrate that there is a strong tendency for certain types of responses to contain continuations, as in the case of the most frequent ones. The most frequent one, *i.e.*, *confirming,* contains 29.4 % of all responses with continuations in the whole corpus, whereas the second most frequent one, *agreeing*, contains 19.7 %.

The continuations that occur in YYRs are of interest to the linguist due to the fact that they frequently contain recurrent word-combinations, or patterns of words and phrases which operate in connecting discourse. Perhaps the most easily perceived type of recurrent word-combinations that occur in YYRs are independent phrases, which are often used by speakers when responding. In section 6.6 we shall, therefore, examine the independent responses which occur in a corpus of *yeah* and *yes* responses.

6.6- The distribution of the independent responses in the corpus

The term *independent responses*, as already noted, has been adopted in the present study to refer to instances of *yeah* or *yes* alone and independent phrases, *i.e.*, short, semi-fixed lexical phrases containing *yeah* or *yes* in the corpus. We shall include instances of *yeah* or *yes* alone among the figures of independent responses, in order to contrast the incidence of such responses with those containing continuations.

As already mentioned in the previous chapters, the valid YYRs in the corpus included instances of *yeah* or *yes* alone which operated as floor returners (*cf*. LEVINSON, 1983) or backchannels (TOTTIE, 1991 *apud* YNGVE, 1970). Backchannels are defined by Tottie (*ibid*., p. 255) as "the sounds (and gestures) made in conversation by the current non-speaker, which grease the wheels of conversation but constitute no claim to take over the turn." Tottie (*ibid*.) also presents some examples of backchannels, which may include vocalizations (*m, mhm, uh[h]-uh*), *yes*, *yeah*, *yes quite*, *surely*, *I see* and *that's true*, among others. They have several functions, "which normally occur simultaneously" and are emphasized by different researchers (TOTTIE, *ibid*., p. 256). These functions are the "supportive function" (when they signal understanding and agreement) and the "regulatory function" (when they are used to encourage the speaker to continue his/her turn). Tottie, however, stresses the supportive function, since the author believes that "a basic component of agreement is normally present, otherwise the non-speaker would not refrain from claiming the turn for any length of time"

(TOTTIE, *ibid*., p. 257). An important aspect of Tottie's discussions is the fact that determining what is to be regarded as a backchannel and what constitutes a turn is a difficult issue. However, the author presents his criteria for determining what is a backchannel, *i.e.*, "backchannel status can be determined only on the basis of the following utterance"; if a phrase uttered by S1 provokes a response from S2, it is a turn and not a backchannel (*cf*. TOTTIE, *ibid*., p. 261).

Therefore, in the present study, we could identify some utterances which were more explicitly used as backchannels (or floor returners, as already mentioned), since they did not provoke responses from S2. In such cases, S2, in fact, proceeded in producing the same utterance after hearing *yeah* or *yes*. Therefore instances of this type were classified as floor returners, or backchannels. However, when there were doubts as to whether single instances of *yeah* or *yes* were backchannels or not, we opted for the illocutionary values which seemed most plausible (in given contexts provided by the co-text) in our tentative classification of illocutionary values).

Thus, instances of *yeah* or *yes* alone which did not operate as floor returners have been classified as instances of independent responses and include instances of repetitions or combinations of *yeah* and *yes*, for example. These were frequently accompanied in the corpus by vocalisations of hesitation and/or pauses, which were not considered as words in the classification.

One type of vocalisation, namely *mm*, however, has already been described in the literature as a word since it is a very frequent item in spoken English. *Mm (*and its variant *Mhm* in the corpus) has been described by Svartvik *et al* (1982 *apud* TOTTIE, 1991, p. 255) as one of the most frequent 'words' in British English conversation. *The Oxford advanced learner's dictionary online* includes *mm* among its entries and describes it as follows,

> **mm** exclamation
> ■*exclamation* (also **mmm**) the way of writing the sound /m/ that people make to show that they are listening to sb or that they agree, they are thinking, they like sth, they are not sure, etc.: *Mm, I know what you mean.* ◇*Mm, what lovely cake!* ◇*Mmm, I'm not so sure that's a good idea.*

Therefore, in our classification of independent responses we have considered *mm* as a word.

The independent responses in the corpus have, thus, been divided into simple (*yeah* or *yes* only), double (*yes yes, yes yeah, m yes/yeah*) and independent phrases (*yeah/yes* [*that's*] [*all*] *right*, *yes I see, oh yes, yes it is*, and others). They may consist of independent response items (*yes, yeah*), clusters of items (*oh yes yes, yes of course*), incomplete clauses (*yes right*) or full clauses (*yeah, that's right*). It should be noted that they may be used in discourse as independent responses but may also be used as responses in conjunction with other sequences of utterances in responses.

Tables 6.8 and 6.9 present the general distribution of the independent responses in the *yeah* and *yes* corpora.

Table 6.8 – General distribution of the independent responses in the *yeah* corpus

Type	**Incidence of tokens**
Simple	**64**
Yeah	64
Double	**14**
Yeah yeah	7
Yes yeah	7
Independent phrases	**16**
Ah Ah yeah	1
Mm. Oh yeah	1
Oh okay. Yeah	1
Oh right yeah yeah yeah	1
Oh yeah	4
Right yeah	3
That's it yeah	1
That's right. Yeah	1
That's right. Yeah. But	1
That's right. Yeah. Mhm	1
Yes that's right yeah	1
Total	**94**

.

Table 6.9 – General distribution of the independent responses in the *yes* corpus

Type	Incidence of tokens
Simple	**46**
Yes	45
Yes?	1
Double	**12**
Mm yes	1
Yes yes	5
Yes yes yes	2
Yes yes yes yes	1
Yeah yes	1
Yes yeah	2
Independent phrases	**37**
Alas yes	1
Did he yes yes	1
It certainly did yes	1
Oh absolutely. Yes. it is.	1
Oh dear. Yes yes	1
Oh yes	1
Oh yes that's right. Yes	1
Oh yes. Yes yes yes	1
(Proper name) *yes* (Proper name/*Sir*)	3
(Oh) Right. Yes	1
Something like that yes	1
Well yes and no	1
Well. Yes that's great.	1
Well yes yes	1
Yeah. Yes. Yes. Yes that's good. Yes	1
Yes. Definitely.	1
Yes exactly	1
Yes. Yes exactly. Yes quite	1
Yes I know what you mean. Yeah. Yeah. Mm.Mm.Mm	1
Yes I see.	1
Yes I see. No. yes. Oh I see	1
Yes I understand yes	1
Yes [erm] okay	1
Yes of course	1
Yes quite	1
Yes + subject + verb (short answers)	6
(Yes)(Yes)(Yes) Yes that's right. (Yeah)	4
Yes. They did. Yes.Yes. Yes.	2
Yes. Yes. Yes. Yes. Quite yes	1
Total	**95**

Tables 6.8 and 6.9 demonstrate that the general distribution of the independent responses containing *yeah* and *yes* presents some striking differences. Although the total number of independent responses (IRs) in the two corpora is almost identical (with 94 instances of *yeah*-IRs and 95 instances of *yes*-IRs), the incidence of *yeah* alone-IRs is 64, against 46 instances of *yes*-only IRs. However, the most significant difference relates to the number of tokens and types (form) of the independent phrases in the two corpora. The *yeah* corpus contains 12 tokens of independent phrases which belong to 9 different types, whereas the *yes* corpus contains a total of 35 tokens of independent phrases which belong to 27 different types.

6.7 –Independent phrases used in longer *yeah* and *yes* responses

As already mentioned in section 6.6, independent phrases can be used in conversation in longer responses, which we have classified as complex responses. Therefore, it is of interest to investigate the frequency at which independent phrases occur in both independent responses and complex responses. This shall be presented in Tables 6.10 and 6.11.

Table 6.10 – Incidence of independent *yeah* phrases in independent and complex responses

Type	Incidence	Frequency in YhRs with continuations (%)	Frequency in the *yeah* corpus (%)
Independent *yeah* phrases used as independent responses	16	7.1	4.8
Independent *yeah* phrases used in complex responses	27	11.9	8.2
Total			**13.0**

.

Table 6.11 – Incidence of independent *yes* phrases in independent and complex responses			
Type	**Incidence**	**Frequency in YRs with continuations (%)**	**Frequency in the *yes* corpus (%)**
Independent *yes* phrases used as independent responses	36	15.2	9.4
Independent *yes* phrases used in complex responses	20	8.5	5.2
Total			**14.6**

The figures presented in tables 6.10 and 6.11 demonstrate that independent phrases are uttered more frequently as independent responses in the *yes* corpus (9.2 % of instances), and among *yes* responses with continuations in general (14.8 %). However, independent phrases occur more frequently in the *yeah* corpus as parts of complex responses (8.2 %), and, also, among *yeah* responses with continuations in general (11.9 %).

If we consider the frequency of the independent phrases from the point of view of the entire corpora of *yeah* and *yes* responses, we can note that these occur at the frequencies of 13.0% and 14.6%. These figures might indicate that these phrases are used at a reasonably steady frequency (since they appeared in the corpora in the 10 to 15 % range); furthermore, they are instances of the phraseology of *yeah* and *yes*, since independent phrases contain words that co-occur with *yeah* and *yes*.

In section 6.8, we shall investigate the collocates of *yeah* and *yes* in the corpus, in an attempt at identifying any additional patterns of collocation that these words may have.

6.8 – The collocations of the words *yeah* and *yes* in the corpus

In this investigation into the collocations of the words *yeah* and *yes*, all the words which appeared on the right-hand side and left-hand side of *yeah* and *yes* within the responses in which they occurred were identified and quantified. It should be noted that the left-hand collocates under investigation are those which preceded the words *yeah* or *yes* within the responses, *i.e.*, they did not appear in the IMs.

We opted for including collocates that contain more than one word when such collocates consist of recurrent phrases that collocate with these two words. Some examples include *it's*, *there is/are/were*, *I think*, *that's right*, which occupy the first slot in the adjoining

sentence. However, the items within these collocates can be treated individually in any study that aims at precisely investigating their occurrences as collocates. For the purposes of the present study, we have considered it more interesting to present them as slot fillers in the position adjacent to *yeah* or *yes*.

Tables 6.12, 6.13 and 6.14 present the right-hand side collocates of *yeah* and *yes* in the two corpora.

Table 6.12 – Right-hand side collocates of *yeah* in all valid responses in the *yeah* corpus

Position	***Yeah***	**Collocates**	**Number of tokens**
1		*he/she/it/we/you/they/people/somebody/*proper name	23
2		*but*	20
3		*and*	18
4		*I*	13
5		*I mean*	9
		there is/there are/there were	9
6		*it's (it is)*	8
7		*that* (referential subject)	7
		well	7
8		*because/'cos*	6
		what	6
9		*I think* (that-clause)	5
		(that's)(all)right	5
		get/go/[be]careful/carry on	4
		mm(mm)	4
11		*no/not*	3
		oh	3
12		*(of) course*	2
		okay	2
		that's right/you're right	2
		this (subject)	2
		lovely/good	2
		although	2
		(that's) true	2
13		*I see*	1
		the	1
		except	1
		if	1
		(Total number of types: 51)	168

.

Table 6.13 – Right-hand side collocates of *yes* in all valid responses in the *yes* corpus			
Position	***Yes***	**Collocates**	**Number of tokens**
1		*he/she/it/we/you/they/people /*proper name	35
2		*I*	21
3		*but*	16
4		*and*	14
5		*because/'cos*	12
		it's (it is)	12
		that's right	12
6		*well*	9
7		*so*	7
		that (referential subject)	7
8		*oh*	6
9		*I mean*	5
10		*I think* (that-clause)	4
		there is/are/were	4
		(all) right	4
		the	4
11		*a/an*	3
		I see	3
		what	3
12		*don't / put* (imperative verbs)	2
		mm	2
		sure	2
		quite	2
13		*actually*	1
		certainly	1
		definitely	1
		exactly	1
		no/not	1
		okay	1
		of course	1
		please	1
		this (subject)	1
		which (referential)	1
		(Total number of types:47)	199

Tables 6.12 and 6.13 demonstrate that *yeah* and *yes* collocate to the right-hand side with a different number of words. *Yeah* collocates with 51 different types of words or phrases, whereas *yes* collocates with 47 different types. Among the right-hand collocates of *yeah*, 47 types occurred in the corpus at an incidence of more than once, whereas the right-hand collocates of *yes occurred* more than once in 37 types. This might indicate that these types tend to be recurrent words that frequently combine with *yeah* and *yes*. The most frequent

collocates of *yeah* and *yes*, which appear at the top of the lists in the two tables, indicate that there are strong collocations of *yeah* and *yes* with specific words.

These most frequent collocates of *yeah* and *yes* are personal pronouns and nouns, which operate as subjects of the adjoining clauses. The same is true of the second most frequent collocate of *yes*, the pronoun *I*, which is the fourth most frequent collocate of *yeah*.

The pronoun *I* has been investigated separately from the other personal pronouns for two reasons: firstly, since it is used in reference to the same speaker who utters the response, it might occur at a higher frequency in relation to those pronouns which refer to other persons; secondly, it appears in the *Cobuild frequency count (top 113 forms)* (*cf*. SINCLAIR, p. 143) as the most frequent pronoun in the entire Bank of English corpus, in the 8th position. However, table 6.14 demonstrates that reference to persons other than the first person singular we more frequent in the two corpora of affirmative responses.

It should be noted that the instances of *I* which were treated separately from the other pernonal pronouns did not include the uses of *I* in the phrases *I see*, *I mean*, and *I think*, which were investigated separately, as two-word phrases. These phrases are used either as discourse markers or, as in the case of *I think*, with *that*-clauses, whose subject is usually other than *I*.

It is interesting to note that the second and third most frequent collocates of *yeah*, which are also the third and fourth most frequent collocates of *yes*, are the conjunctions *but* and *and*. These conjunctions operate in the initial slot of the adjoining clauses.

The fifth most frequent collocate of *yeah*, namely, *I mean*, occurred in 9 instances, and represented the ninth most frequent collocate of *yes*, with 5 instances. The fifth most frequent collocate of *yes* was *because*, with 12 instances. *Because* was the sixth most frequent collocate of *yeah*, with 6 instances.

In the sixth position among the collocates of *yeah*, the forms *there is/are* had 9 occurrences, and they appeared in the tenth position among the collocates of *yes*, with 4 occurrences.

The independent phrase *(that's) right* occurred 12 times in the *yes* corpus, as the fifth most frequent collocate, against 5 occurrences in the *yeah* corpus, in the ninth position.

So was seventh most frequent collocate of *yes*, with 7 instances; however, it did not collocate with *yeah* in the corpus.

These figures, in addition to the others presented in tables 6.12 and 6.13, demonstrate that there are some significant differences between the two corpora, in addition to some similarities. These similarities are compared in table 6.14 below, which describes the right-

hand collocates of both *yeah* and *yes*. They are linguistic items which appeared in both corpora, and have been summed up in one comparative table.

.

Table 6.14 – Right-hand side collocates of both *yeah* and *yes* in all valid responses in the corpus

Position	*Yeah* and *Yes*	Collocates	Number of tokens
1		*he/she/it/we/you/they/people /*proper name	58
2		*but*	36
3		*I*	34
4		*and*	32
5		*it's (it is)*	20
6		*because/'cos*	18
7		*so*	16
		well	16
8		*I mean*	14
		that's right	14
9		*there is/are/were*	13
10		*I think* (that-clause)	9
		oh	9
		(all) right	9
		what	9
11		imperative verbs *(yeah+get/go; yes+don't /put)*	6
		mm	6
12		*the*	4
13		*I see*	3
		no/not	3
		(Total number of types: 34)	329

Table 6.14 demonstrates that *yeah* and *yes* were used in the corpus with thirty-five types of right-hand side collocates. Both have strong collocational patterns, especially as in the case of the ten most frequent positions. All other instances occurred more than once, which indicates a tendency for the co-occurrence of these items too.

Pronouns and nouns referring to people represent the first and third most frequent items, totalling 92 instances.

The second and third most frequent items, the conjunctions *but* and *and*, confirm the tendency for cohesive items to collocate with *yeah* and *yes*, which had been demonstrated through the instances of pronouns and nouns in the previous paragraph.

The only instance of a non-cohesive item that collocated with *yeah* and *yes* is represented by the verb *there to be*. *There is/are/were* are the ninth most frequent collocates

of *yeah* and *yes*. All other collocates, however, are cohesive items, which include, among others, elliptical forms (such as *oh*, imperative verbs, *mm*, *no* and *not*.

Tables 6.15, 6.16 and 6.17 present the left-hand side collocates of *yeah* and *yes* which were uttered by the same speaker (*i.e.*, S2), and which preceded *yeah* and *yes* within the responses.

.

Table 6.15 – Left-hand side collocates of *yeah* in all valid responses in the *yeah* corpus			
Position	**Collocates**	***Yeah***	**Number of tokens**
1	*Oh*		12
	(oh)(that's) right		12
2	*Well*		5
3	*Mm*		3
4	*I see*		2
	(that's) it		2
	Okay		2
5	*[Ah] ah*		1
	because		1
	good		1
	Hopefully		1
	nice		1
	quite well		1
	really		1
	so		1
	sometimes		1
	terrible		1
	will		1
	you		1
	you know		1
	(Total number of types: 24)		51

.

Table 6.16 – Left-hand side collocates of *yes* in all valid responses in the *yes* corpus			
Position	**Collocates**	***Yes***	**Number of tokens**
1	*Oh*		27
2	*(oh)(that's) right*		8
3	*Well*		7
4	pronoun/proper name		4
5	*but*		3
	mm		3
6	*Hello*		2
	I think		2
	that		2
7	*absolutely*		1
	Ah ah		1
	Alas		1
	are		1
	did		1
	I think so		1
	it is		1
	Oh dear		1
	Oh my God		1
	of course		1
	said		1
	so		1
	that's it		1
	You see		1
	(Total number of types: 27)		72

.

Table 6.17 – Left-hand side collocates of *yeah* and *yes* in all valid responses in the corpus			
Position	**Collocates**	***Yeah/Yes***	**Number of tokens**
1	*Oh*		39
2	*(oh)(that's) right*		20
3	*Well*		12
4	*mm*		6
5	pronoun/proper name		4
6	*but*		3
	[that's] it		3
7	*[Ah] ah*		2
	so		2
	(Total number of types: 14)		91

The figures for the left-hand side collocates of *yeah* and *yes* demonstrate that they consisted of linguistic items which were also used as right-hand collocates in most cases, except for the instances of phrases such as *Ah ah*, *Hello*, *Alas*, *Oh dear* and *Oh my God*, and the adjectives *nice* and *terrible*, which totalled 9 instances out of 123 instances of left-hand collocates in the two corpora.

The left-hand side collocates occurred much less frequently than the right-hand ones, with 51 instances in the YhC and 72 instances in the YC, thus totalling 123 instances, against 367 instances of right-hand collocates (c*f*. tables 6.12 and 6.13).

Table 6.17 demonstrates that the left-hand side collocates of both *yeah* and *yes* were all cohesive items. Furthermore, especially the three most frequent ones occurred with *yeah* and *yes* in very strong collocational patterns. They are *Oh*, with 39 instances, *(oh) (that's) right*, with 20 instances, and *Well*, with 12 instances. All other instances of collocates in table 6.17 occurred more than once, which indicates their tendency for co-occurrence with *yeah* and *yes*, despite the low frequencies displayed.

In the next section, we shall demonstrate that the right-hand side collocates of *yeah* and *yes*, are, in fact, integral parts of the lexical phrases that co-occur with these two words. The patterns constituted by the sequences of such lexical phrases will be investigated and quantified.

6.9- Patterns of sequences of lexical phrases in *confirming* responses

As already mentioned in section 6.5, the most frequent *yeah* and *yes* responses which contained continuations in the corpus are *confirmings*, which totalled 137 instances (*cf*. tables 6.5 to 6.7). The responses which contained continuations were realised by instances of *yeah* or *yes* which were followed by independent or complex responses, as already mentioned in sections 6.6 and 6.7 (*cf*. tables 6.8 to 6.11).

The continuations in *yeah* responses were investigated in relation to the lexical phrases that constituted them, according to the scheme of analysis proposed by Nattinger and DeCarrico (1992; *cf*. Chapter 4 in the present study). We opted for investigating such sequences of lexical phrases in *yeah* responses only, since *yeah* is more informally used than *yes* in conversation, and since we aimed at investigating the patterns of some specific linguistic items (*cf*. sections 6.10 and 6.11 as well) in a more informal register (which would represent casual, informal conversation).

Yeah confirmings had an incidence of 64 instances in the corpus and the most frequent category of lexical phrases that they contained is type 1b, "conversational purpose". Conversational purpose lexical phrases (henceforth referred to as 1b) co-occurred in all instances of *yeah confirmings* (henceforth referred to as *confirmings*) at least twice, except for 2 instances of 1b followed by type 1a ("conversational maintenance") only, and one instance of 1b followed by type 3 ("discourse devices").

The sequences of lexical phrases encountered in the continuations to *yeah* in *confirmings*, consisted of up to six lexical phrases in direct sequence (*i.e.*, with no intervening items). This maximum number of LPs in a sequence was partly due to the fact that the number of words in the span (on each line of the concordances) is limited; on certain occasions, S2 was interrupted by another speaker, or S2 produced shorter sequences of LPs. Therefore, there were sequences of LPs of various lengths. Since the length of the sequences is not the focus of the investigations, we shall not present any figures about the number of LPs in each sequence. We shall, therefore, concentrate on the main patterns of sequences encountered.

Some examples of the sequences of LPs that were encountered include:

#28 3+1b+3+1b+3 (fluency device + *confirming* + logical connector + *justifying* + fluency device)

#95 1b+3+1b+3 (*confirming* + logical connector + *justifying* + logical connector)

#153 1b+1b (*confirming* + *confirming*)

#268 1b+1b+1b+1b+1b (*confirming* + *confirming* + *confirming* + *informing* + *giving directions*)

#35 1b+3+1b+3+1b+3 (*confirming* + summarizer + *confirming* + fluency device + *giving opinion* + evaluator)

#344 1b+1a+1a+1b (*confirming* + *clarifying* + *clarifying* + *informing*)

The observable patterns of LPs in *confirming* responses were classified according to the combinations of *confirming* LPs with the adjoining categories of LPs (that occurred in each response). Therefore, the combinations of *confirming* LPs occurred in eight different types of sequences (in decreasing order of frequency): a) *confirming* and *informing* (1b+1b); b) two or three *confirmings* (or *endorsing confirmation*) and another type 1b (1b1b[1b]+1b); c) *confirmings* only or *confirming* and *endorsing* only (1b+1b+[1b]+[1b]+[1b]); d) *confirmings* and various other LPs (1b+X); e) *confirming* and *informing* and another type 1b (1b+1b+1b); f) *confirming* and *justifying* (1b+1b); g) *confirming* and *clarifying* (1b+1a); and h) *confirming* and *giving opinion* (1b+1b).

The most frequent sequence, *confirming* and *informing* (1b+1b) LPs occurred at an incidence of 21 cases. *Informing* LPs were followed by other types of LPs (type 3 in 12 instances, and type 2, "necessary topic", in 2 instances). *Informing* was also followed by other instances of *informing* (in two instances).

The second most frequent sequences occurred 16 times. They consisted of two or three *confirmings* (or *endorsing confirmation*) and another type 1b LP (1b+1b [+1b]+1b). The two or three sequences of *confirmings* were followed by *clarifying* (2 instances), *giving opinion* (2 instances), and *requesting* (one instance). The sequences of two or three instances of *confirmings* and *informings* were followed by *giving directions*, *informing* and *seeking confirmation* (one instance each); they also contained 9 instances of type 3 LPs and initial, medial or final position.

The third most frequent sequence of LPs is *confirmings* only or *confirming* and *endorsing* only (1b+1b+[1b+1b+1b]), with 6 instances. *Endorsings* were included in this group since they were endorsements to the previous *confirmings*, *i.e*., they conveyed a similar illocutionary value to that of *confirmings*. These sequences consisted of sequences of *confirmings* only (5 instances), and one instance of *confirmings* and *endorsings* interspersed. There was only one instance of type 3 LP in used in conjunction with the sequences of *confirmings*.

The fourth most frequent sequence consisted of *confirmings* followed by 5 different 1b LPs, namely *fact*, *hypothesising*, *requesting information*, *comparing* and *seeking confirmation*. These were accompanied by 2 instances of type 3 LPs and one instance of type 1 LP (*giving opinion*).

In fifth position, there were three distinct frequent sequences. One of them is the sequence *confirming* and *informing* followed by another type 1b LP (1b+1b+1b), which occurred in 4 instances. The other type 1b that occurred in these sequences were *clarifying*, *giving opinion*, *hypothesising* and *seeking confirmation*. These were accompanied by two type 3 LPs and 2 type 1b LPs (*giving opinion* and *clarifying*).

Another sequence that occurred 4 times is *confirming - justifying* (1b+1b). These 4 instances co-occurred with type 3 LPs in 3 instances, and with one instance of type 1a; one of these sequences was followed by type 1b *requesting information*.

The sequence *confirming – clarifying* (1b+1a) also occurred 4 times; none of these sequences was followed by another type 1 or 2 LP, and one of them co-occurred with 2 instances of type 3 LPs.

The sixth most frequent sequence of LPs, *confirming - giving opinion* (1b+1b) occurred 3 times in the corpus. Each of these patterns was followed by *giving opinion, endorsing* and *justifying*, which are all type 1b LPs. One of them also included one instance of type 1b LP *giving opinion*) and one instance of type 3 LP in the sequence; one of these sequences contained one instance of type 3 LP.

The identification of these sequences of LPs in the continuations of the *confirming yeah* responses demonstrates that lexical phrases occurred in specific patterns in these responses. These patterns had different frequencies of occurrence in the corpus: the most frequent one had an incidence of 21 instances, the second one had 16 instances, and the third one appeared in 6 cases (out of a total of 64 instances). Therefore, it can be argued that when responding affirmatively through the use of *yeah* in confirmations, speakers tend to produce patterns of sequences of LPs which operate in reinforcing the confirmation that was initially signalled (through *yeah*) or expanding the topic to a related area, such as justifying the action, clarifying some previous information, or giving opinion, among others.

In section 6.10 we will investigate other linguistic items that operate in the structuring of connected spoken discourse, namely verbs and their tenses. We will identify the most frequent verbs and tenses that occurred in the corpus and their patterns of occurrence.

6.10- The most frequent verbs in *confirming yeah* responses

The most frequent verbs that were encountered in the corpus of *yeah confirmings* are presented in table 6.18 below. We have included in this table only those verbs which occurred more than once in the corpus.

Table 6.18- The most frequent verb forms in *confirming yeah* responses		
Position	**Verb form**	**Number of instances**
1	*is*	27
2	*was*	5
	(I) mean	5
3	*know*	4
4	*think* (with one negative form *don't think*)	3
	used to (with one negative form *didn't use*)	3
5	*are*	2
	do	2
	go/goes	2
	have	2
	have/has been	2
	have got (one elliptical)	2
	have heard	2
	remember	2
	there is	2
	went	2
	(Others, 1 instance each)	(27)

The most frequent verb form encountered in the continuations of *confirming* YhRs, *is*, appeared in 27 instances, against the second most frequent ones, *was* and *mean* (in the phrase *I mean*), which had 5 instances each. The third most frequent one is the verb know, with four instances, and the fourth most frequent ones are *used to* and *think*, with three instances each (including one negative form each). The other verbs presented in table 6.18 occurred twice each, and there were also 27 instances of other verbs which occurred only once in *confirming* YhRs.

The most frequent verb forms on the list indicate that *yeah* tends to occur in patterns of lexical preference with some verbs in *confirming* YhRs. *Is*, *was*, *mean* and *know* are among the most preferred lexical items that tend to co-occur with *yeah* in instances of *confirming*.

An investigation of the most frequent verb tenses used in the continuations of *confirming* YhRs demonstrated that the preferred tenses are: the present simple, with 63 instances, the past simple, with 21 instances, and the present perfect, with 3 instances. Four other tenses (the second conditional "*would* + infinitive", the third conditional "*would have*", the past perfect, and "*going to*" future forms) occurred only once each. Therefore, the present

simple, the past simple and the present perfect can be described as the most preferred verb (syntactic) structures that tend to co-occur with *yeah* in *confirming* YhRs.

In section 6.11 we will investigate the patterns of co-occurrence of cohesive items in *confirming* YhRs.

6.11- Items of cohesion in *confirming yeah* responses

In our investigations into the items of cohesion that operate within the continuations of *confirming yeah* responses, we analysed all instances of individual words that occurred in them. Therefore, each word is considered as one linguistic item in the general count. The only exceptions are verb phrases with more than one word, the cases of reiteration and rephrasing, and the expressions *at the same time* and *as bad as*, which count as one linguistic item each.

We identified a total of 524 linguistic items within the continuations of *yeah confirmings*. Among them, there were 300 cohesive items. These cohesive items belonged to the following types:

1- reference: 132 instances;
2- ellipsis: 96 instances (including all instances of *yeah*);
3- conjunction: 52 instances;
4- substitution: 7 instances (including 5 instances of rephrasing); and
5- lexical cohesion (reiteration): 6 instances.

After identifying the distribution of the different types of cohesion among the continuations of *yeah confirmings*, we analysed their patterns of occurrence in relation to the distance from the node *yeah*. Table 6.19 includes the cohesive items that occurred before and after the node *yeah*, and their respective distances (up to the eighth position to the right; there were, however, some few instances of cohesive items in the ninth position, which are not shown here).

The findings can be summarised in table 6.19 below.

.

Table 6.19- Distribution of cohesive items in *yeah confirming* responses (distances from the node *yeah*)												
Type of cohesive item	3	2	1	***yeah***	1	2	3	4	5	6	7	8
conjunction	-	1	3	-	20	4	6	5	1	2	2	4
reference	4	1	1	-	23	24	18	13	12	14	11	6
ellipsis	-	1	5	79	2	3	-	2	1	1	1	-
lexical	-	-	1	-	1	1	-	-	-	-	-	-
substitution	-	-	2	-	2	1	1	1	-	-	-	-
Total	4	3	12	79	48	33	25	21	14	17	14	10

Table 6.19 shows that the continuations of *yeah* in *confirming* responses contained cohesive items in all slots to the right- and to the left-hand side of *yeah*. The highest incidence of cohesive items can be observed in the two slots on the right-hand side of *yeah*, which also contain all the five general types of cohesive items. The first slot on the left-hand side also contains all the five types of cohesive items. From the third to the eighth slot on the right-hand side of *yeah*, there is a high incidence of cohesive items as well, with a greater number of instances of conjunction and reference items. Lexical cohesion is the type with the lowest rate of incidence, however.

The figures about the incidence and distribution of cohesive items demonstrate that *yeah* responses, here represented by instances of *confirmings* and their continuations, have highly cohesive contents. The slots which were not occupied by reference items consisted of verbs, adjectives and adverbials. Cohesive items, therefore, can be considered as collocates of *yeah*, and table 6.19 demonstrates that items of reference and conjunctions, for example, have strong collocational patterns with *yeah*.

In section 6.11 below, we shall present a discussion of the findings obtained in the previous sections.

6.12- Discussion of findings

The analyses of the *yeah* and *yes* responses in this chapter have allowed for the identification of significant facts about the occurrences of these responses in the corpus. The initial analyses revealed that the two corpora contained responses which were originated by

different initiation moves, the *yeah* corpus contained 25 types of IMs among which there were 9 types that did not occur in the *yes* corpus. The *yes* corpus contained 26 types of IMs; among these, there were 10 types which did not occur in the *yeah* corpus. Thus, there was a total of 19 differing illocutionary values of IMs in the two corpora, which demonstrates that the two corpora were different with respect to the illocutionary values of the IMs and YYRs they contained.

Another important finding relates to the frequency at which the IMs were used in the two corpora. The three most frequent IMs in the two corpora were the same, namely *informing*, *giving opinion* and *requesting information*. These three most frequent IMs account for the majority of the IMs in the two corpora, *i.e*, 450 instances out of a total of 697 IMs in the two corpora (*cf*. table 6.3).

Although the four most frequent illocutionary values of the YYRs did not occur at the same frequencies (the decreasing order of frequency was *confirming, agreeing, acknowledging* and *informing* in the YhC, and *acknowledging, confirming, agreeing* and *informing* in the YC), the differences in the number of their occurrences were very small; these four illocutionary values accounted for the majority of responses in both corpora. *Confirming* occurred in 197 instances, *acknowledging* in 192 instances, *agreeing* in 144 instances and *informing* in 76 instances. They account for 533 responses out of a total of 713 in the two corpora.

Furthermore, the most frequent IMs generated the majority of the most frequent responses: *informing* IMs generated 161 *acknowledgements* and 54 *confirmings*, *giving opinion* generated 115 *agreeings*, and *requesting information* generated 73 *informings*, among other responses (*cf*. tables 6.1 and 6.2).

These figures confirm hypothesis 1 in the present study (*cf*. Chapter 1), since the YYRs in the corpus were realised by speakers in prevailing patterns of illocutionary values. The incidence of the most frequent values of *yeah* and *yes* demonstrates that YYRs tend to occur more frequently in utterances of specific illocutionary values. Furthermore, they were generated more frequently by IMs that were used in the performance of specific illocutionary values.

The combinations of the most frequent IMs and the YYRs they generated illustrate the occurrence of preferred sequences and sequence types, in accordance with the discussions into this topic which were developed in Chapter 2 (*cf*. section 2.4). This is well illustrated, for example, through the case of the most frequent IM, namely *informing*. *Informing* IMs were mostly realised through affirmative form sentences. The most frequent YYRs they generated

(*acknowledging* and *confirming*) thus represented the most preferred sequences to them, constituting patterns of sequences (which can be contrasted, for example, with the least preferred sequences).

In another phase of the analyses, we investigated the valid YYRs in the corpus which contained continuations, since we aimed at identifying any collocational patterns of the words *yeah* and *yes*, and of the phrases within the continuations. The analyses revealed that *yeah* and *yes* frequently co-occur with specific words which can belong to independent phrases in independent responses or which belong to complex responses, thus connecting parts of the discourse. We identified the combinations of *yeah* in 16 types of independent phrases (*cf*. table 6.8) and combinations of *yes* in 37 different types of independent phrases (*cf*. table 6.9), which represents a significant difference in the number of occurrences of the two words in the corpus.

We also discovered that the frequency of the independent *yeah* phrases used in independent responses was 4.8%, as opposed to 9.8% of cases of independent *yes* phrases used as independent responses. However, both *yeah* and *yes* independent phrases occurred in the two corpora at similar frequencies (13.0% and 14.6% respectively), since they were also used in complex responses (11.9% and 8.5% respectively).

The analyses of the collocates of *yeah* and *yes* revealed that *yeah* collocated to the right-hand side with 51 types of phrases or words, at an incidence of 168 cases. *Yes* collocated with 47 different types of phrases or words to the right-hand side, at an incidence of 199 cases. The phrases or words that collocated with both *yeah* and *yes* totalled 34 types. Therefore, we could verify that *yeah* and *yes* did not collocate with the same phrases or words in all instances, since they only had 34 collocates in common in the corpus.

The left-hand side collocates within the responses were much less frequent, as we already expected. *Yeah* had 24 different types of left-hand side collocates, at an incidence of 51 cases; *yes* had 27 different types of left-hand side collocates, at an incidence of 72 instances. Both *yeah* and *yes* had 14 different left-hand side collocates in common, which totalled 91 instances. Therefore, the left-hand side collocates of both *yeah* and *yes* were not the same in all instances; in fact, the number of left-hand side collocates they shared (14) is nearly half the number of the left-hand side collocates of *yes* only (27).

These findings confirm hypothesis 2 in the present study (*cf*. Chapter 1), since the figures demonstrate that *yeah* and *yes* collocate more frequently with some phrases and words. The findings also demonstrate that the collocational patterns of *yeah* and *yes* are not identical.

The analyses of the occurrences of lexical phrases within the *confirming yeah* responses, which were the most frequent *yeah* responses which contained continuations, revealed that there were eight major patterns of sequences of LPs within them. The sequences contained, in most cases, LPs which belonged to category 1b, *i.e.*, conversational purpose (*cf*. the scheme for analysis in section 5.4). Co-occurrences of category 1b LPs permeated all the patterns which were encountered, except for one pattern, which consisted of the combination 1b +1a (conversational maintenance). The patterns that were identified (*cf*. section 6.9) contained, in some cases, instances of category 3 LPs (*i.e.*, "items that connect the meaning and structure of the discourse", according to the scheme for classification), but not all of them did. However, it should be noted that in our classification of the patterns we prioritised the combinations of categories 1a, 1b and 2, since the propositional content of the utterances within the responses belonged to these categories. Category 3 LPs, due to their basic characteristic of connecting the structure of the discourse, did not consist of complete utterances and, thus, had a secondary role in the major patterns we identified.

However, although category 3 LPs are defined as items that connect the meaning and structure of the discourse, we verified that the connection of meaning and structure was not realised by these LPs only. Further analyses into the cohesive items within the responses and the lexical and syntactic preferences of *yeah* revealed that these elements played a decisive role in connecting meaning and structure. This fact reinforced the importance of the patterns of lexical phrases which we identified, since the latter were densely constituted by instances of such elements.

The three most frequent sequences of LPs occurred at 21, 16 and 6 instances (*cf*. section 6.9). These figures demonstrate that *confirming* YhRs are realised in patterns of LPs that occur at different frequencies, thus confirming hypothesis 5 in the present study in relation to YhRs (*cf*. Chapter 1). The identification of the eight patterns of sequences of LPs and of the highest incidence of some of these patterns indicates that *yeah* and *yes* responses in general may consist of patterns of sequences of LPs which occur at different frequencies. These findings also confirm the arguments posed by Sinclair (1991; 1987), Nattinger and DeCarrico (1992), Widdowson (1991), Hunston (2002), McCarthy (2001; 1991), and other researchers who have argued in favour of a phraseological view of language.

This is further corroborated by the findings obtained from another component of the analyses, namely, the investigations into the occurrences of verb forms and tenses. The figures demonstrated that *yeah* co-occurred with the verb form *is* in 27 instances, out of a total of 94 verb forms; the second most frequent ones, namely *was* and *mean* (in the phrase *I mean*)

occurred 5 times each. They were followed by the form *know*, with 4 instances; other verb forms followed these ones (*cf*, table 6.18).

The most frequent verb tense that co-occurred with *yeah* was the present simple, with 63 instances (only one negative form); the second most frequent one was the past simple, with 21 instances (one negative form); the third most frequent one was the present perfect, which occurred three times only (*cf*. section 6.10 for a complete description).

As a result of these findings, it could be argued that *confirming* YhRs frequently occur in patterns of preference for the syntactic structures of the present and past simple, the present perfect and the other verb tenses encountered in the corpus. In these cases, there was a strong lexical preference for the verb forms *is, was* and *mean*, in addition to the other less frequent verbs encountered. All the other verbs identified in table 6.18 occurred more than once, which indicates their tendency to co-occur with *yeah* in *confirmings*.

These findings confirm hypothesis 3 in relation to *yeah* responses. By analogy, we could argue that *yes* responses undergo a similar process of displaying lexical and syntactic selection preferences.

The final part of the analyses dealt with the items of cohesion in *confirming* YhRs. Among the 524 linguistic items that exist in these responses, 300 items consisted of elements of cohesion. Ellipsis was the most frequent type of cohesion, with 96 instances. This was largely due to all the instances of *yeah* which appeared in the responses. *Yeah* is described by Halliday and Hasan (1976) as an elliptical item of cohesion. However, there were also a few instances of elliptical subject and verb in the corpus.

The second most frequent type of cohesion was conjunction, with 52 instances, followed by substitution, which had 7 instances, and lexical cohesion (reiteration), with 6 instances. In addition, we discovered that reference and conjunction were the types of cohesive items that occurred most frequently at the distances of one or two words from the node *yeah* (*cf*. table 6.19). Reference was also very frequent at the distances of 3 to 8 words located on the right-hand side of *yeah*.

Therefore, hypothesis 4 could be confirmed in relation to YhRs since the frequencies and location of the cohesive items they contained demonstrated that such items occurred in very strong collocational patterns which combined the type of cohesive item and its distance from the word yeah. By analogy, we could argue that *yes* responses are also heavily dependent on items of cohesion.

The findings demonstrate that YYRs are realised in the language in patterns of collocation of phrases and words which operate at the different levels of the organisation of

discourse. The pragmatic, formal and phraseological patterns encountered in the analyses reveal that YYRs are instances of connected spoken discourse which consist of strings of utterances that are inter-connected supra- and intra-sententially. The patterns that operate between both the different types of IMs and responses, and within the responses themselves, confirm the phraseological characteristic the responses investigated.

In the next chapter we will present the final conclusions.

7- CONCLUSION

7.1- The contributions of the present study to corpus linguistics, pragmatics and phraseology

The findings obtained in this study confirm our arguments in favour of an analysis of *yeah* and *yes* responses as instances of spoken connected discourse. As noted in the Introduction (Chapter 1), traditional studies in linguistics have been based on data which are often simplified due to the fact that most of it is based on intuition or written language. On the other hand, studies which focus on naturally-occurring spoken discourse had, for a long time, been placed on a secondary level, partly due to the fact that conversational language had not been systematically described in the literature.

However, the recent contributions of corpus linguistics have allowed for the use of corpora in investigations into the properties of spoken discourse. Studies which focus on natural connected discourse can, for instance, enable the analyst to investigate both extra- and intra-sentential phenomena which have been traditionally ignored, thus combining the contributions of corpus linguistics, pragmatics and studies into phraseology.

The analyses that were developed in the present study confirmed our initial hypotheses about the co-occurrence of various and diverse phenomena which we believed to be frequent characteristics of responses in conversation. As already mentioned in Chapter 5, the choice for the nodes *yeah* and *yes* in the corpus was due to the fact that these words were accompanied by the initiation moves and the continuations of the responses, which allowed us to investigate the various items presented in the hypotheses. The initiation moves and continuations, as the co-text, provided the major contextual clues for the identification and classification of the illocutionary values of the utterances in the corpus (*cf*. Chapters 2, 5 and 6).

The nodes *yeah* and *yes* were analysed separately for the purposes of clarity, and the analyses revealed some differences about the uses of the two words, the responses and the initiation moves. We discovered that *yeah* and *yes* responses were originated by different sets

of IMs, since not all of them had the same illocutionary values. The number of differing illocutionary values amounted to 19 categories. Another important finding was the fact that the most frequent IMs in the two corpora were the same ones and occurred in strong frequencies, since they totalled 450 instances out of the total of 697 IMs.

The most frequent illocutionary values of *yeah* and *yes* were the same ones, too, although in different order of frequency. The two corpora contained, however, different sets of illocutionary values of *yeah* and *yes*, and there was a total number of 6 differing illocutionary values. We could verify, therefore, that *yeah* and *yes* responses occur in the language in preferred patterns of sequences in association with the initiation moves that originate them. In addition, there were preferred types of initiation moves which were followed by *yeah* and *yes* responses, which demonstrates that their occurrences are not random in the language. Another interesting finding relates to the fact that *yeah* and *yes* responses are generated, in most cases, by statements, rather than questions, which would be counter-intuitive, since traditional descriptions usually present responses which are uttered after questions.

The analyses also revealed that, within the responses which contained continuations, the words *yeah* and *yeah* collocated with specific words and phrases in patterns of preference. They also collocated, for example, with different sets of independent phrases at different frequencies. *Yeah* collocated with 16 types of independent phrases, whereas *yes* collocated with 37 different types of independent phrases. The frequencies of use of independent phrases in independent responses differed as well. As for the collocations with words, we identified a higher incidence of right-hand collocates with *yes*; we could also discover that *yeah* and *yes* did not collocate with identical sets of words to the right- or left-hand sides. Consequently, these findings demonstrated that *yeah* and *yes* occur in the language in preferred patterns of association with specific words and phrases, and that the patterns are not identical for both.

In order to narrow down our focus of analysis, we investigated the most frequent *yeah* response in the corpus, namely *confirming*. *Confirming* YhRs were chosen, firstly, because *confirming* was the most frequent illocutionary value of both *yeah* and *yes* responses with continuations; and secondly, because we opted for the more informal register represented by the use of *yeah*. *Confirmings* were submitted to investigations into the occurrences of lexical phrases, verbs and cohesive items. The results demonstrated that the lexical phrases contained in *confirmings* occurred in observable patterns of preferred sequences, the most frequent ones accounting for the majority of cases. The same was true in relation to the occurrences of verb forms and tenses, which presented higher frequencies of specific forms and tenses, thus

demonstrating that the responses contained preferred lexical and syntactic choices. As for cohesion, two types were used by speakers much more frequently than the others, and the proximal slots they occupied in relation to the node *yeah* within the responses also showed their strong collocational properties.

These findings demonstrate that studies into the collocational, phraseological and pragmatic characteristics of spoken connected discourse can produce a variety of insights, which are only partly exemplified through the analyses. The combination of pragmatics and corpus studies, for example, can be a valuable resource in the development of further in-depth analyses of responses of the type investigated in the present study. It is hoped, therefore, that the contributions of the present study to the fields of corpus linguistics, pragmatics and phraseology can be of interest to researchers and language teachers.

7.2- Research limitations

The investigations that were developed in the present study represented the initial stages of the analyses that are possible through the use of computerised corpora. Large corpora provide empirical evidence which requires human observation for the identification of significant linguistic patterns. The use of concordance pages allows for the observation of patterns and for decision-making in relation to the use of additional resources. Thus, further computer assistance, through the use of specially-designed software, is necessary if the focus of the investigations is shifted towards, for example, very frequent words that collocate with the word in the node. Another example is the case of the examination of the lexical bundles (*cf*. Chapter 4) that exist in the corpus. In the *yeah* and *yes* corpora it would be interesting to identify the most frequent lexical bundles, but that would require some specially-designed software.Therefore, there were some technological limitations, which did not, however, impair the work that was developed in consonance with the original aims.

A limitation of a different nature is represented by the identification of the illocutionary values of the utterances in the corpus, which lacked contextual evidence. Although the units of analysis and the intuitive judgements can be criticised, they were theoretically supported (*cf*. Chapters 2 and 3); furthermore, intuition has been described by Sinclair (1991) and Hunston (1999) as intrinsic to the descriptions of corpora.

7.3- Further research

The findings derived from the present study could be further investigated in various aspects. For instance, further investigations into responses which are originated by questions might present some evidence about the constraints involved in their production. Another example, the responses which are originated by statement IMs are very frequent but contain different frequency distributions. Some less frequent IMs could be further investigated through the compilation of corpora with more instances of such IMs. The same could be done in relation to the responses that they originate.

The very frequent responses, however, could be further investigated through the use of sub-corpora which would display two lines of information for each utterance in the concordance pages. These could be analysed through the use of software which would allow for the identification of distal collocates, patterns, lexical phrases, and lexical bundles, for example. Further research about the occurrences of specific linguistic items, such as adjectives and adverbs would also be of interest for the identification of lexical preferences within the responses.

To sum up, the findings in the present study can be further developed in various aspects. In most cases, the use of the computer will be of paramount importance.

7.4- Pedagogic implications

The pedagogic implications of the present study are related to the emphasis we have placed on the phraseological characteristic of language. We have argued throughout this study that a phraseological view of language is of importance for a better understanding of the workings of the language. Therefore, studies into phraseology can largely contribute to a shift in paradigm in relation to the production of classroom materials and reference books, and the development of new approaches to teaching and syllabus design (*cf*. the discussions in Chapter 3).

These new developments would, hopefully, aid learners in the acquisition of native-like proficiency. As noted by Kjellmer (1991) the native speaker has "an automatic command of substantial portions of speech and uses his pauses to plan one or more thought units ahead. In building his utterances he makes use of large prefabricated sections." However, the learner,

> having automated few collocations, continually has to create structures that he can only hope will be acceptable to native speakers; he, too, will have to plan his thought units, but we can assume that his pauses are to a great extent used for decision-making at this fairly trivial word-structure level. His building material is individual bricks rather than prefabricated sections.
> So even if he is not diffident, uncertain or hesitant he will inevitably be hampered in his progress, and his output will often seem contrived or downwright unacceptable to native ears. Analogous phenomena can be observed in his written output.
> (KJELLMER *apud* AIJMER and ALTENBERG 1991, p.124.)

In the same respect, Widdowson argues that communicative competence is "knowing a stock of partially pre-assembled patterns, formulaic frameworks, and a kit of rules, so to speak, and being able to apply the rules to make whatever adjustments are necessary according to contextual demands" (WIDDOWSON 1989, p. 135)

The author cites Sinclair, who argues that

> Now that we have the means to observe samples of language which must be fairly close to representative samples, the clear messages are:
> a) We are teaching English in ignorance of a vast amount of basic fact. (...)
> b) The categories and methods we use to describe English are not appropriate to the new material. We shall need to overhaul our descriptive systems.
> c) Since our view of the language will change profoundly, we must expect substantial influence on the specification of syllabuses, design of materials and choice of method. (SINCLAIR 1985, *apud* WIDDOWSON, 1990., p. 76).

However, according to Widdowson (*ibid.*), any decisions about the language that is pedagogically presented should always be "informed by an understanding of what theoretical assumptions underlie the different descriptions. […] In other words, and understanding of theory is needed for evaluation."

Therefore, the present study, with its emphasis on the phraseological characteristic of language, aims at presenting additional evidence about the underlying principles that operate in connected discourse. We hope that the insights provided in the discussions can be of utility to further studies into connected spoken discourse.

8- BIBLIOGRAPHY

AIJMER, Karin. *Conversational routines in English – convention and creativity.* London: Longman, 1996. 251 P.

AIJMER, Karin; ALTENBERG, Bengt. *English corpus linguistics*. London: Longman, 1991. 338 p.

ALTENBERG, Bengt. On the phraseology of spoken English: the evidence of recurrent word-combinations. In: COWIE, Anthony Paul (Ed.) *Phraseology – Theory, analysis, and applications*. Oxford: OUP, 1998. p. 55-78.

AUSTIN, John L. *How to do things with words*. Oxford: OUP, 1962 [1986]. 176p.

BACH, K.; HARNISH, R.; M. *Linguistic communication and speech acts.* Cambridge, MA: MIT Press, 1979.

BALLY, C. *Linguistique générale et linguistique française.* Berne: Franck, 1932. In: KERBRAT-ORECCHIONI, Catherine. *Os atos de linguagem no discurso. Teoria e funcionamento.* Niterói: EDUFF, 2005. 220 p.

BENVENISTE, E. *Problèmes de linguistique génerale*. Paris: Gallimard, 1966. t. I. In: KERBRAT-ORECCHIONI, Catherine. *Os atos de linguagem no discurso. Teoria e funcionamento.* Niterói: EDUFF, 2005. 220 p.

BERNARDINI, S. *Competence, capacity, corpora*. Bologna: CLUEB, 2000. In: HUNSTON, S. *Corpora in applied linguistics*. Cambridge: CUP, 2002. 241 p.

BIBER, Douglas; CONRAD, S.; REPPEN, R. *Corpus linguistics: investigating language structure and use*. Cambridge: CUP, 1998. 300p.

BIBER, Douglas; JOHANSSON, Stig; LEECH, Geoffrey; CONRAD, Susan; FINEGAN, Edward. *Longman grammar of spoken and written English*. London: Longman, 1999. 1204p.

BLUM-KULKA, Shoshana. Discourse pragmatics. In: VAN DIJK, Teun. *Discourse as social interaction*. London: SAGE, 1997. 324 p.

BLUM-KULKA, Shoshana, HOUSE, Juliane, KASPER, Gabriele. (Eds.) Cross cultural pragmatics: requests and apologies. In: *Advances in discourse processes*, volume XXI. New Jersey: ABLEX Publishing Corp., 1984 [1989]. 302 p.

BROWN, Penelope, LEVINSON, Stephen. Universals in language usage: politeness phenomena. In: Goody, E.N. (Ed.), *Questions and politeness: strategies in social interaction.* (Cambridge Papers in Social Anthropology, 8.) Cambridge: CUP, 1978 [1987] . 394 p.

BÜHLER, s/n. 1933. In: KERBRAT-ORECCHIONI, Catherine. *Os atos de linguagem no discurso. Teoria e funcionamento.* Niterói: EDUFF, 2005. 220 p.

The Cambridge Advanced Learner's Dictionary Online. Available at: < http://dictionary.cambridge.org>. Last access: 3rd April, 2006

CELCE-MURCIA, M. Grammar pedagogy in second and foreign language teaching. TESOL Quarterly 25 (3), 1991, p. 459-480.

COLE, P.; MORGAN, J.L. (Eds.) *Syntax and semantics 3: speech acts.* New York: Academic Press, 1975. 215 p.

Collins Dictionary Online. Available at: http://www.collins.co.uk/wordexchange/Sections/DicSrchRsult.aspx?word=yes. Last access: 3rd April, 2006

COMRIE, B. *Aspect: an introduction to the study of verbal aspect and related problems.* Cambridge: CUP, 1976. 126 p.

COWIE, Anthony Paul. Stable and creative aspects of vocabulary use. In: CARTER, Ronald; McCARTHY, Michael. *Vocabulary and language teaching*. London: Longman, p. 126-139.

_____. (Ed.) *Phraseology – Theory, analysis, and applications*. Oxford: OUP, 1998. 258 p.

CRYSTAL, David. *The Cambridge encyclopaedia of language*. Cambridge: CUP, 1987. 472 p.

CRYSTAL, David. *The Cambridge encyclopaedia of the English language*. Cambridge: CUP, 1995. 489 p.

CUNNINGHAM, Sarah; MOOR, Peter. *Cutting edge: pre-intermediate.* Harlow: Longman, 2001. 168 p.

DAVIDSON, J. Subsequent versions of invitations, offers, requests and proposals dealing with potential or actual rejection. In: Atkinson, J.; Heritage, J. *Structures of social action. Cambridge: CUP* (in press). In: LEVINSON, Stephen C. *Pragmatics*. Cambridge: CUP, 1983 [1987]. 420 p.

DE COCK, Sylvie. A recent word combination approach to the study of formulae in the speech of native and non-native speakers of English. International Journal of Corpus Linguistics 3, 1998, p. 59-80. In: HUNSTON, S. *Corpora in applied linguistics*. Cambridge: CUP, 2002. 241 p.

DE COCK, Sylvie; GRANGER, Sylviane; LEECH, Geoffrey; McENERY, Tony. An automated approach to the phrasicon of EFL learners. In: GRANGER, Sylviane (Ed.) *Learner English on computer*. London: Longman, 1998, 228 p.

DÖRNYEI, Z., THURRELL, S. Strategic competence and how to teach it. *ELT Journal*, 45 (1), 1991: 16-23

DREW, Paul; HOLT, Elizabeth. Idiomatic expressions and their role in the organization of topic transition in conversation. In: EVERAERT, Martin et al. *Idioms: structural and psychological perspectives*. New Jersey: Lawrence Erlbaum Associates, 1995. 329 p.

ELLIS, Rod. *The study of second language acquisition*. Oxford: OUP, 1994. 824 p.

Encarta Dictionary. Available at:
<http://encarta.msn.com/dictionary>. Last access: 3rd April, 2006-09-11

EVERAERT, Martin et al. *Idioms: structural and psychological perspectives*. New Jersey: Lawrence Erlbaum Associates, 1995. 329 p.

FARR, F., ROMMETVEIT, R. The communicative act: an epilogue to mutualities in dialogue. In: Marcová, I., Grauman, C., Foppa, K. (Eds.) Mutualities in dialogue. Cambridge: CUP, 1995, p. 264-274. In: McCARTHY, Michael. *Issues in applied linguistics*. Cambridge: CUP, 2001. 175p.

FILLMORE, C. 'Corpus linguistics' or 'computer-aided armchair linguistics'. In: SVARTVIK, J. (Ed.) *Directions in corpus linguistics*. Berlin and New York: Mouton de Gruyter, 1992. p. 35-60.

FIRTH, J.R. *Papers in linguistics 1934-1951*. London: OUP, 1957. In: WIDDOWSON, Henry G. *Aspects of language teaching*. Oxford: OUP, 1990. 213p.

_____. (unidentified) 1953. In: STUBBS, Michael. *Discourse analysis: the sociolinguistic analysis of natural language*. Oxford: Basil Blackwell, 1983. 250 p.

FRANCIS, W. N.; KUCERA, H. *Frequency analysis of English usage*. Boston: Houghton Mifflin Company, 1982

GARDINER, A. H. *Langage et acte de langage. Aux sources de la pragmatique*. Lille: Presse universiteire de Lille, 1989. In: KERBRAT-ORECCHIONI, Catherine. *Os atos de linguagem no discurso. Teoria e funcionamento.* Niterói: EDUFF, 2005. 220 p.

GLEDHILL, C. Collocation and genre analysis: the phraseology of grammatical items in cancer research abstracts and articles. *Zeitschrift für Anglistik und Amerikanistik*, v. 43, 1995. p. 11-29.

GOODY, E.N. (Ed.), *Questions and politeness: strategies in social interaction*. (Cambridge Papers in Social Anthropology, 8.) Cambridge: CUP, 1978 [1987]. 394 p.

GRANGER, Sylviane. Prefabricated patterns in advanced EFL writing: collocations and formulae. In: COWIE, Anthony Paul (Ed.) *Phraseology – Theory, analysis, and applications*. Oxford: OUP, 1998. p. 145-160.

_____. The computer learner corpus: a versatile new source of data for SLA research. In: GRANGER, Sylviane. (Ed.) *Learner English on computer*. London: Longman, 1998. p. 3-18.

GRICE, H.P. Logic in conversation. In: COLE, P. and MORGAN, J.L. (Eds.) *Syntax and semantics 3: speech acts.* New York: Academic Press, 1975, pp. 41-58

HALLIDAY, M., HASAN, R. *Cohesion in English.* London: Longman, 1976

HORNBY, A.S.; COWIE, A.P. (Eds) *Oxford advanced learner's dictionary of current English*. Oxford: OUP, 1987 [1974]. 1041 p.

HUGHES, G., McCARTHY, Michael. From sentence to discourse: discourse grammar and English language teaching. TESOL Quarterly, 32 (2), p. 263-287.

HUNSTON, S. *Corpora in applied linguistics*. Cambridge: CUP, 2002. 241 p.

HUNSTON, S., FRANCIS, G. *Pattern grammar: a corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins, 2000. 289p.

HYMES, Dell. On communicative competence. In: PRIDE, J.B.; HOLMES, J. (Eds.) *Sociolinguistics: selected readings*. Hammondsworth: Penguin books, 1972.

JAKOBSON, R. *Essais de linguistique générale.* Paris: Minuit, 1963. In: KERBRAT-ORECCHIONI, Catherine. *Os atos de linguagem no discurso. Teoria e funcionamento.* Niterói: EDUFF, 2005. 220 p.

JEFFERSON, G. s/n. 1980 and 1984. In: BROWN, P., LEVINSON, S.C. Universals in language usage: politeness phenomena. In: Goody, E.N. (Ed.), *Questions and politeness: strategies in social interaction*. (Cambridge Papers in Social Anthropology, 8.) Cambridge: CUP, 1978 [1987]. 394 p.

KERBRAT-ORECCHIONI, Catherine. *Os atos de linguagem no discurso. Teoria e funcionamento.* Niterói: EDUFF, 2005. 220 p.

KJELLMER, Göran. A mint of phrases. In: AIJMER, Karin; ALTENBERG, Bengt. *English corpus linguistics*. London: Longman, 1991. 338 p.

LABOV, W.; FANSHEL, D. s/n, 1977. In: BROWN, P., LEVINSON, S.C. Universals in language usage: politeness phenomena. In: Goody, E.N. (Ed.), *Questions and politeness:*

strategies in social interaction. (Cambridge Papers in Social Anthropology, 8.) Cambridge: CUP, 1978 [1987]. 394 p.

LEECH, Geoffrey. Corpora and theories of linguistic performance. In: SVARTVIK, J. (Ed.) *Directions in corpus linguistics*. Berlin and New York: Mouton de Gruyter, 1992. p. 105-122.

_____. *Principles of pragmatics*. London: Longman, 1983 [1984]. 249p.

LEVINSON, Stephen C. *Pragmatics*. Cambridge: CUP, 1983 [1987]. 420p.

LEWIS, Michael. Implications of a lexical view of language. In: WILLIS, Jane; WILLIS, Dave (Eds.). *Challenge and change in language teaching*. Oxford: Macmillan, 1996. p. 10-16.

_____ (Ed.) *Teaching collocation*. Hove: LTP, 2000

Longman dictionary of contemporary English online, updated edition. Available at: <http://www.ldoceonline.com/>. Last access: 3rd April, 2006

MARTINS, Helena. Sobre a estabilidade do significado em Wittgenstein. Veredas, Juiz de Fora: Editora UFJF, v. 7, p. 19-42, jan./dez. 2000.

McCARTHY, Michael. *Discourse analysis for language teachers.* Cambridge: CUP, 1991. 213 p.

_____. *Issues in applied linguistics*. Cambridge: CUP, 2001. 175p.

McCARTHY, Michael, CARTER, Ronald. *Language as discourse: perspectives for language teaching*. London: Longman, 1994. 230 p.

_____. Written and spoken vocabulary. In: SCHMITT, Norbert; McCARTHY, Michael. (Eds.) *Vocabulary: description, acquisition and pedagogy*. Cambridge: CUP, 1997. 383 p.

Merriam-Webster Online dictionary. Available at:
< http://www.m-w.com>. Last access: 3rd April, 2006

MEY, Jakob. *Pragmatics: an introduction*. Oxford: Blackwell, 1993. In: VAN DIJK, Teun. *Discourse as social interaction*. London: SAGE, 1997. 324 p.

MINDT, D. *An empirical grammar of the English verb system*. Berlin: Cornelsen, 2000.

MOON, Rosamund. *Fixed expressions and idioms in English*. Oxford: Clarendon Press, 1998. 338 p. (a)

_____. Frequencies and forms of phrasal lexemes in English. In: COWIE, A.P. (Ed.) *Phraseology – Theory, analysis, and applications*. Oxford: OUP, 1998 (b)

_____. The analysis of meaning. In: SINCLAIR, John. *Looking up – An account of the COBUILD Project in lexical computing*. London: Collins, 1987. 182 p.

_____. Vocabulary connections: multi-word items in English. In: SCHMITT, Norbert; MCCARTHY, Michael. (Eds.) *Vocabulary: description, acquisition and pedagogy*. Cambridge: CUP, 1997, 383 p.

MORRIS, C. *Foundations of the theory of signs*. Foundations of the unity of science: towards an international encyclopaedia of unified science, I, 2. Chicago: The university of Chicago press, 1938. In: VERSCHUEREN, Jef. *Understanding pragmatics*. London: Arnold, 1999. 295 p.

NATION, P., WARING, R. Vocabulary size, text coverage and word lists. In: SCHMITT, Norbert; MCCARTHY, Michael. (Eds.) *Vocabulary: description, acquisition and pedagogy*. Cambridge: CUP, 1997. 383 p.

NATTINGER, James; De CARRICO, Jeanette. *Lexical phrases and language teaching*. Oxford: OUP, 1992. 218 p.

OXENDEN, Clive; SELIGSON, Paul. *English file 1: student's book*. Oxford: OUP, 1996. 144p.

Oxford Advanced Learner's Dictionary Online, 7th Edition. Available at: <http://www.oup.com/oald-bin/web_getald7index1a.pl>. Accessed on 3rd April, 2006.

PARRET, H. et al. *Le langage en contexte.* Amsterdam: John Benjamins, 1980. In: VERSCHUEREN, Jef. *Understanding pragmatics*. London: Arnold, 1999. 295 p.

PAWLEY, Andrew; Syder, Frances. Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In: RICHARDS, Jack, Schmidt (Eds.) Language and communication. London: Longman, 1983. 276 p.

PETERS, A. language learning strategies: does the whole equal the sum of the parts? In: Language, v. 53 (3), 1977. p. 560-573.

POMERANTZ, A. Compliment responses: notes on the co-operation of multiple constraints. In: Schenkein, J. (Ed.) *Studies in the organization of conversational interaction*. New York: Academic Press, 1978. p. 79-112.

_____. Why compliments get rejected: a sequential analysis of compliment responses. In: BROWN, P., LEVINSON, S.C. Universals in language usage: politeness phenomena. In:

Goody, E.N. (ed.), *Questions and politeness: strategies in social interaction*. (Cambridge Papers in Social Anthropology, 8.) Cambridge: CUP, 1978 [1979] / 1987[1990]. 394 p.

RENOUF, Antoinette. Moving on. In: SINCLAIR, John (Editor). *Looking up*: *an account of the COBUILD Project in lexical computing*. London: Collins ELT, 1987. 182 p.

RICHARDS, J. *The context of language teaching*. London: CUP, 1985. 228 p.

RICHARDS, Jack; PLATT, John; PLATT, Heidi. *Longman dictionary of language teaching and applied linguistics.* London: Longman, 1992. 423 p.

ROSALDO, M. The things we do with words: Ilongot speech acts. In: COLE, P.; MORGAN, J. (Eds.) *Syntax and semantics 9: pragmatics*. New York: Academic press, 1978. p.261-281.

SEARLE, John. Indirect speech acts. In: COLE, P.; MORGAN, J. (Eds.) *Syntax and semantics 3: speech acts.* New York: Academic Press, 1975. p. 59-82.

_____. *Speech acts: an essay in the philosophy of language.* Cambridge: CUP, 1969.

SCHEGLOFF, E. A., JEFFERSON, G., SACKS, H. The preference for self-correction in the organization of repair in conversation. *Language*, v. 53, 1977. p. 361-82. In: LEVINSON, Stephen.C.. *Pragmatics*. Cambridge: CUP, 1983 [1987]. 420 p.

SCHMIDT, Robert. The role of consciousness in second second language learning. *Applied linguistics* **11**, 1990, p. 129-158. In: HUNSTON, S. *Corpora in applied linguistics*. Cambridge: CUP, 2002. 241 p.

SCHMIDT, S. J. (Ed.) Pragmatik I: Interdisziplinäre Beiträge zur erforschung der sprachlichen Kommunikation. München: Fink, 1974. In: VERSCHUEREN, Jef. *Understanding pragmatics*. London: Arnold, 1999. 295 p.

SCHMITT, Norbert; MCCARTHY, Michael. (Eds.) *Vocabulary: description, acquisition and pedagogy*. Cambridge: CUP, 1997. 383 p.

SILVEIRA PEDRO, Verônica. *A corpus analysis of yes-responses in English: the cases of agreeing and informing*. Unpublished M.A. dissertation. Niterói: UFF, 1999. 124 p.

SINCLAIR, John. A way with common words. In: H. HASSELGARD and S. OKSEFJELL (Eds.) *Out of corpora: Studies in honour of Stig Johansson*. Amsterdam: Rodopi, 1999, p. 157-179.

_____. *Corpus, concordance, collocation.* Oxford: OUP, 1991. 179 p.

_____. (a) (Editor). *Collins Cobuild English language dictionary*. London: Collins ELT, 1987. 1703 p.

_____. (b) (Editor). *Looking up: an account of the COBUILD Project in lexical computing*. London: Collins ELT, 1987. 182 p.

_____. (c). Grammar in the dictionary. In: SINCLAIR, J. (Ed.) *Looking up: an account of the COBUILD Project in lexical computing*. London: Collins ELT, 1987

_____. (d) The nature of the evidence. In: SINCLAIR, J. (Ed.) *Looking up*: *an account of the COBUILD Project in lexical computing*. London: Collins ELT, 1987.

_____. Selected issues. In: QUIRK, RANDOLPH; Widdowson, Henry (Eds.) *English in the world*. Cambridge: CUP, 1985. In: WIDDOWSON, Henry. *Aspects of language teaching*. Oxford: OUP, 1990. 213p.

SINCLAIR, John *et al*. *Collins COBUILD English grammar*. London: HarperCollins, 1990. 486 p.

SINCLAIR, John; RENOUF, Antoinette. A lexical syllabus for language learning. In: CARTER, Ronald; McCARTHY, Michael (Eds.) *Vocabulary and language teaching*. London: Longman.

STENSTRÖM, A. B. 1990. Lexical items peculiar to spoken discourse. In: SVARTVIK, J. (Ed.) *The London-Lund corpus of spoken English*. Lund: Lund University Press, p. 137-75.

STUBBS, Michael. *Discourse analysis: the sociolinguistic analysis of natural language*. Oxford: Basil Blackwell, 1983. 250 p.

SUMMERS, D. Computer lexicography: the importance of representativeness in relation to frequency. In: THOMAS, Jenny; SHORT, Michael (Eds). *Using corpora for language research. London: Longman*, 1996. In: HUNSTON, S. *Corpora in applied linguistics*. Cambridge: CUP, 2002. 241 p.

SVARTVIK, J.; EEG-OLOFSSON, M; FORSHEDEN, O. ORESTRÖM, B; THAVENIUS,C. *Survey of spoken English. Report on research 1975-81*. Lund: Lund University Press, Lund, 1982.

THOMAS, Jenny. Cross-cultural pragmatic failure. Applied linguistics v. 4, p. 91-112. In: ELLIS, Rod. *The study of second language acquisition*. Oxford: OUP, 1994. 824 p.

TOGNINI-BONELLI, E. The role of corpus evidence in linguistic theory and description. PhD thesis, University of Birmingham. Published as *Corpus Theory and Practice*. Birmingham: twc. In: HUNSTON, S., FRANCIS, G. *Pattern grammar: a corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins, 2000. 289p.

TOTTIE, Gunnel. Conversational style in British and American English: the case of backchannels. In: AIJMER, Karin; ALTENBERG, Bengt. *English corpus linguistics*. London: Longman, 1991. p. 254-271.

VERSCHUEREN, Jef. *Understanding pragmatics*. London: Arnold, 1999. 295 p.

WANNER, L. (Ed.) *Lexical functions in lexicography and natural language processing*. Amsterdam: John Benjamins, 1996.

WIDDOWSON, Henry G. *Aspects of language teaching*. Oxford: OUP, 1990. 213p.

_____. Knowledge of language and ability for use. Applied linguistics v. 10. Oxford: OUP, 1989. p. 128-37.

_____. *Aspects of language teaching*. Oxford: OUP, 1990. 213 p.

WIERZBICKA, A. Different cultures, different languages, different speech acts. Journal of pragmatics, v. 9. 1985, p. 145-163.

WILLIS, Dave. *Rules, patterns and words*. Cambridge: CUP, 2003. 238p.

_____. The idiom principle: a pedagogic approach. Paper read at the University of Birmingham, November 1998. In: HUNSTON, S. *Corpora in applied linguistics*. Cambridge: CUP, 2002. 241 p.

_____. *The lexical syllabus*. London: Collins ELT, 1990. 139 p.

WILLIS, Jane; WILLIS, Dave (Eds.). *Challenge and change in language teaching*. Oxford: Macmillan, 1996. 186 p.

Wikipedia Dicionary. Available at:
< http://en.wiktionary.org/wiki/yes> Last access: 3rd April, 2006

WITTGENSTEIN, Ludwig. *Investigações filosóficas*. São Paulo: Nova cultural, 2000. 207 p.

YNGVE, V. *On getting a word in edgewise*. Papers from the sixth regional meeting of the Chicago linguistic society. Chicago: University of Chicago, 1970. p 567-577.

YORIO, C. Idiomaticity as an indicator of second language proficiency. In: HYLTENSTAM, K.; OBLER, L.K. (Eds.) *Bilingualism across the lifespan*. Cambridge: CUP, 1989. p. 55-72.

YULE, George. *Pragmatics*. Oxford: OUP, 1996. 138 p.

9- APPENDICES

9.1- Form for classifying individual IMs and *yeah* or *yes* responses

No.	Initiation move	Response	Lexical phrases [ ]
			Quantity [ ]

IM:……

……..

R:……

……

Response			
Yes alone [ ]	Short answer [ ]	Clause (s) [ ]	Phrase(s) only [ ]
Lexical phrase(s)			

Type(s):
1- Social interaction markers
1a- Conversational maintenance
1b- Conversational purpose
2- Necessary topics
3- Discourse devices

9.2- Form for each illocutionary value of the IMs (in the *yeah* or *yes* corpora) and the responses it generated

IM:	No.	Node	Continuation(s)

9.3- Form for each illocutionary value of the IMs (in the *yeah* and *yes* corpora) and the responses it generated (arranged according to their illocutionary values and their totals)

IM (illocutionary value):	**Illocutionary values of the responses**	**Total**
Total number of instances:	……...	……
	……...	……
	……...	……
	……...	……
	……...	……
	……...	……
	……...	……

9.4- Form for each illocutionary value of *yeah* or *yes* and its continuations

	Illocutionary values of the continuations:	**Total:**
Illocutionary value of *YEAH* [*or YES*]:	…….....	……..
	…….....	……..
……	…….....	……..
	…….....	……..
	…….....	……..
Total no.: ……	…….....	……..
	…….....	……..
	…….....	……..

9.5- Form for the identification of the collocational patterns of the cohesive items in *confirming yeah* responses

Nº	**Left-hand collocates**				**Right-hand collocates**							
	3	2	1	***yeah***	1	2	3	4	5	6	7	8
		C	R	E	*E.g.* R	R	C					
					R	R	L					
					R	E	S					
Type of cohesive item												
conjunction												
reference												
ellipsis												
lexical												
substitution												
Total												

Symbols: C = conjunction; R = reference; E = ellipsis; L = lexical cohesion; S = substitution

9.6 – Affirmative responses encountered in a pilot study (data collected from conversations in British films and interviews)

Type 1- short answers

Form: (*Yes* +) personal pronoun + *am / is / are / was / were / do / does / did / can / could / have / has / had / will / would*

Examples:

	I/	am /	was /	do /							
		.....									
(Yes,)	you / we / they	are /	were /	do /	did/	can/	could/	have/	had/	will/	would
		.....									
	he / she / it	is /	was /	does /							

Type 2- Miscellaneous (different forms)

A bit
A little
Absolutely
And you shall
Anything
Aye
Brilliant
Certainly
Don't worry
Exactly
Excellent
Fair enough
Fine / That's fine / Seems fine
Grand
I guess so
I know
I see
I won't!
I think so
I think we could manage that
I suppose yes ...
I suppose so
In a way
Looks like it
Maybe
Me too
No, I suppose not
Nor do/does/did I/he/she

No, I guess not
Of course not
OK
Okey dokey
Only when
I believe so
I thought so
I'm afraid (so/yes)
I'll do my best
I understand
I give you my word
I remember
I agree
I'll be there
I thought he/she/it might
I appreciate that
(It's) Great
I give you my word
I understand
If you absolutely insist
Indeed
It doesn't matter
I'll do my best
It's a crime
It's a shame
Just one / a little / a small piece
Might be right
No problem
Perfect
Please do / (Yes) Please
Right / That's right / All right / That's alright / You're right / You must be right
Same here
Splendid
So do/does/did I/he/she
So it/he/she is
Sure
That's a good reason
That's it
The sooner, the better
This is it
True
Very good / Good/Good thinking /(That's a) Good idea
Very nice
Very well
Well, yes
Well, the (subject) did
Why not?
Why yes
Yes, here I am

9.7 – Tokens of the different types of initiation moves and *yeah* responses

9.7.1- Statement initiation moves

#131 He always just stands out in my mind

YEAH. And did anybody know did (Informing – Acknowledging information)

#3 like to keep my lemons for Pancake Day.

YEAH oh I made a lot yesterday. (Giving opinion – Agreeing)

#52 We'll see what our listeners think but thanks for

YEAH probably others have got different opinions. The woman that (Decision – Complying)

9.7.2- Question initiation moves

#224 Shall I wait down there like for you?

YEAH. (Requesting agreement – Agreeing)

#2 Is the pub packed?

Er YEAH there's quite a lot of people in the function room (Requesting information – Informing)

#223 [I used to work there.] Did you yeah.

YEAH. Well I used to work between them out there and (Requesting confirmation – Confirming)

9.7.3- Command initiation moves

22 We've got to make some more up here anyway. Have a last look around.

YEAH. Are there enough there (Requesting action - complying)

#202 Fold the table there's always a #

YEAH, there's a little crack (Requesting action - complying)

#17 Separate it and just shake it off.

YEAH. You can have a freebie FX to use it.

(Giving instructions – Complying)

9.8 - Tokens of the different illocutionary values of *yeah* responses

#173 Oh a hundred and six pounds

I'll have YEAH. (Informing – **Accepting**)

#123 that I was banned from a football ground.

YEAH I accept that what I'm saying is part of the trouble I (Informing – **Accepting**)

#10 I have no objection to the treatment of 'laboured'.

YEAH. (Informing – **Acknowledging**)

#59 I ought to put these books back in the bookcase.

YEAH. Just put them on top nice and tidy. (Necessity – **Acknowledging**)

#224 Shall I wait down there like for you?

YEAH. (Requesting agreement – **Agreeing**)

#3 like to keep my lemons for Pancake Day.

YEAH oh I made a lot yesterday. (Giving opinion – **Agreeing**)

22 We've got to make some more up here anyway. Have a last look around.

YEAH. Are there enough there (Requesting action - **Complying**)

#202 Fold the table there's always a #

YEAH, there's a little crack (Requesting action -C**omplying**)

#223 [I used to work there.] Did you yeah.

YEAH. Well I used to work between them out there and (Requesting confirmation – **Confirming**)

#33 thirteen. Erm so er yep there's something wrong here.

YEAH. What I'm saying is the main (Fact - **Confirming**)

#154 Yes I was one of the broadcasters. Yes that's right

YEAH (Informing – **Endorsing**)

#39 Yes I'll go along with that.

YEAH. Good. Er you always do I (Decision – **Endorsing**)

#319 centre to go in.

I think that's really good.YEAH. (Informing – **Giving opinion**)

#19 Oh yes I'd be interested to would you?

YEAH it would be interesting to read it. Because it does (Requesting opinion - **Giving opinion**)

#2 Is the pub packed?

Er YEAH there's quite a lot of people in the function room (Requesting information – **Informing**)

#15 Oh really?

YEAH just turned twenty it was her birthday on Monday. (Requesting confirmation – **Confirming**)

#274 (No initiation move)

Coke? And I'll say YEAH I'm not going to say No. I just say yeah. Erm Mx? I mean does he sort of (**Reporting speech**)

9.9 – Tokens of the different types of initiation moves and *yes* responses

9.9.1- Statement initiation moves

#54 reason is er they've got no fixed address

YES they have. They're there and the council (Informing – Contradicting)

#365 They meet fairly regularly and

YES Oh yes. They got lots of interests in (Informing – Contradicting)

#71 That's a good one. Students get ripe.

YES they have to (Giving opinion – agreeing)

9.9.2- Question initiation moves

#16 Do you ever play that hedgehog game?

YES actually now now this is something I MX and I took it home (Requesting information – Informing)

#1 Yes what the sort of Gorbals one?

YES exactly. Er although I come from the Gorbals (Requesting confirmation – Confirming)

#4 Indeed this is the whole point isn't it?

Well YES but that wasn't fifty years ago you (Requesting opinion – Giving opinion)

9.9.3- Command intiation moves

#25 do this and let me know what reply you get as well.

YES I shall do # (Advising – Accepting)

#45 Get it in writing as well get things in writing.

YES. YES. YEAH we have been trying to do that (Advising – Accepting)

#70 Just find them FX. They'll be there.

YES. (Advising – Accepting)

9.9.4- Summons

#94 Hello.

YES MX?

#198 Hello John.

YES FX.

#30 Hello.

YES sir.

9.10 - Tokens of the different illocutionary values of *yes* responses

#25 do this and let me know what reply you get as well.

YES I shall do # (Advising – **Accepting**)

#45 Get it in writing as well get things in writing.

YES. YES. YEAH we have been trying to do that (Advising – **Accepting**)

#9 It's our market.

YES Cos I presume by Ireland he means Northern (Informing – **Acknowledging**)

#10 Well they're the same text

YES YEAH but the format (Fact – **Acknowledging**)

#69 It's like good Wimbledon game this. Oh yes.

Oh YES YES. I do remember it (Giving opinion – **Agreeing**)

#71 That's a good one. Students get ripe.

YES they have to (Giving opinion – **Agreeing**)

#1 Yes what the sort of Gorbals one?

YES exactly. Er although I come from the Gorbals (Requesting confirmation – **Confirming**)

you listen in colour don't you?

Oh YES.YES. I think it's very very good. (Requesting confirmation – **Confirming**)

#54 reason is er they've got no fixed address

YES they have. They're there and the council (Informing – **Contradicting**)

#365 They meet fairly regularly and

YES Oh yes. They got lots of interests in

(Informing – **Contradicting**)

#139 It's up near the Hagley Road.

Oh of course it is. YES it's not as convenient as the Vale site. (Informing – **Endorsing**)

#445 (No initiation move)

impression from the DoE was that

YES it was left very much for MX to get on with (**Endorsing**)

#4 Indeed this is the whole point isn't it?

Well YES but that wasn't fifty years ago you (Requesting opinion – **Giving opinion**)

#59 Do you think it does?

Oh YES garlic in things is wonderful. It's beautiful flavour. (Requesting opinion – **Giving opinion**)

#8 can I say anything about a possible visit in

YES. Definitely. (Requesting permission – **Giving permission**)

#391 Can I come round and cry on your shoulder?

YES of course. (Requesting permission – **Giving permission**)

#280 (Hello John. How are you?) All right. And you MX?

YES thanks. Now I think we're running away with the thing here (Greeting – **Greeting**)

#16 Do you ever play that hedgehog game?

YES actually now now this is something I MX and I took it home (Requesting information – **Informing**)

#24 Is it expensive?

Erm it is a bit. YES. (Requesting information – **Informing**)

#252 (No initiation move)

been asked to run in the hundred on the Saturday so I said YES to Middlesex you see to go down to Southampton and I (**Reporting speech**)

#351 (No initiation move)

he looked at FX and she said Oh YES well it is. And in the end he looked that stupid you know (**Reporting speech**)

#213 used to train when everybody else had gone home

Did he YES YES. (Informing - **Requesting confirmation**)

#100 Yeah you're right.

YES? (Agreeing - **Requesting confirmation**)

#198 Hello John.

YES FX. (Summoning – **Responding to summons**)

#30 Hello.

YES sir. (Summoning – **Responding to summons**)

#338 (No initiation move)

I always get that the wrong way round but YES so when you go to Susse so it's only two hours on the train (**Shifting topic**)

#423 (No initiation move)

they did er allow us to get on with the job. So YES DoE and predatory ambitions. No. I don't think so. (**Shifting topic**)