UNITED KINGDOM · CHINA · MALAYSIA

The University of Nottingham

Knight, Dawn (2009) A multi-modal corpus approach to the analysis of backchanneling behaviour. PhD thesis, University of Nottingham.

Access from the University of Nottingham repository:
http://eprints.nottingham.ac.uk/10786/1/Phd_Knight.D.pdf

Copyright and reuse:

For more information, please contact eprints@nottingham.ac.uk

A multi-modal corpus approach to the analysis of backchanneling behaviour

A thesis submitted to The University of Nottingham
for the degree of Doctor of Philosophy in
Applied Linguistics

June 2009

Dawn Knight, MA, BA
School of English Studies
The University of Nottingham

Contents

Abstract

Current methodologies in corpus linguistics have revolutionised the way we look at language. They allow us to make objective observations about written and spoken language in use. However, most corpora are limited in scope because they are unable to capture language and communication beyond the word. This is problematic given that interaction is in fact multi-modal, as meaning is constructed through the interplay of text, gesture and prosody; a combination of verbal and non-verbal characteristics.

This thesis outlines, then utilises, a multi-modal approach to corpus linguistics, and examines how such can be used to facilitate our explorations of backchanneling phenomena in conversation, such as gestural and verbal signals of active listenership. Backchannels have been seen as being highly conventionalised, they differ considerably in form, function, interlocutor and location (in context and co-text). Therefore their relevance at any given time in a given conversation is highly conditional.

The thesis provides an in-depth investigation of the use of, and the relationship between, spoken and non-verbal forms of this behaviour, focusing on a particular sub-set of gestural forms: head nods. This investigation is undertaken by analysing the patterned use of specific forms and functions of backchannels within and across sentence boundaries, as evidenced in a five-hour sub-corpus of dyadic multi-modal conversational episodes, taken from the Nottingham Multi-Modal Corpus (NMMC).

The results from this investigation reveal 22 key findings regarding the collaborative and cooperative nature of backchannels, which function to both support and extend what is already known about such behaviours. Using

these findings, the thesis presents an adapted pragmatic-functional linguistic coding matrix for the classification and examination of backchanneling phenomena. This fuses the different, dynamic properties of spoken and non-verbal forms of this behaviour into a single, integrated conceptual model, in order to provide the foundations, a theoretical point-of-entry, for future research of this nature.

Acknowledgements

List of Figures

Guide to Acronyms, Corpora and Software

Adobe Premiere	Digital editing software developed by Adobe
AL	Applied Linguistics
AMI	Augmented Multi-Party Interaction Corpus
ANVIL	Multimodal annotation and visualisation tool
Atlas.ti	A qualitative-based multi-media analysis tool
AU	Action Units (part of FACS)
.avi	Audio Video Interleaved
BNC	British National Corpus
BoE	Bank of English Corpus
CA	Conversational Analysis
CANCODE	Cambridge and Nottingham Corpus of Discourse in English
CES	Corpus Encoding Standard
CHILDES	Child Language Data Exchange System
CID	Corpus of Interactional Data
CL	Corpus Linguistic
CLAN	Allows for the coding and analysis of text, compatible with the CHILDES corpus/ transcription database
CLAWS	Constituent-Likelihood Automatic Word-Tagging System
CNV	Convergence Token backchannel
CON	Continuer backchannel
Constellations	An 'event based' analysis tool that allows users to synchronise and time align multiple modes of data
CV	Computer Vision
DA	Discourse Analysis
DAMSL	Dialog Act Markup in Several Layers
Diver	Allows users to synchronise, view and play multiple video streams (part of Dynapad)
DReSS	Digital Records for eSocial Science, a 3 year ESRC-funded project based at the University of Nottingham.
DRS	Digital Replay System
DV	Digital Video
Dynapad	A multimodal visualisation (representation) tool
ELAN	A multimedia analysis and representation tool
ER	Engaged Response backchannel
ESRC	Economic and Social Research Council (UK)
EUDICO	European Distributed Corpora Project
FACS	Facial Action Coding Scheme (Ekman and Friesen, 1978)
HamNoSys	Hamburg Notion System (for encoding sign language)
HCI	Human-Computer Interaction
HH	Human-Human Interaction
HMM	Hidden Markov Model
ICSI	International Computer Science Institute (Berkeley, CA)
IFADV	Institute of Phonetic Sciences Free Dialog Video Corpus
IMDI	ISLE Metadata Initiative
I-Observe	An ethnographic data collection and organisation tool
IR	Information Receipt backchannel
ISLE	International Standards for Language Engineering
KAMS	Kernel Based Annealed Mean Shift Algorithm

LCIE	The Limerick Corpus of Irish English
LLC	London-Lund Corpus
LOB	Lancaster-Oslo/ Bergen Corpus
MDCL	Meeting Data Collection Laboratory (held at NIST)
MediaTagger	Mac based software that facilitates the codification of video data at different 'tiers', developed at MPI
MI	Mutual Information
MIBL	Multimodal Instruction-Based Learning Corpus (for Robots)
MM	Multi-Modal
MM4	Multi-Modal Meeting Corpus
MPI	Max Planck Institute
MSC	Mission Survival Corpus
MultiTool	A multimodal transcription and analysis tool
MUMIN	A Nordic Network for MUltiModal Interfaces
MuTra	A multimodal transcription tool
NCeSS	National Centre for eSocial Science
NIMM	Natural Interaction and Multi-Modal Annotation Schemes
NIST	National Institute of Standards and Technology
NITE XML	A workbench of tools that allows for the annotation of natural interactive and multimodal data
NLP	Natural Language Processing
NMMC	The Nottingham Multi-Modal Corpus
NVB	Non-Verbal Behaviour
NVC	Non-Verbal Communication
Nvivo	A tool that supports the alignment and analysis of multiple multi-media data streams
OLAC	Open Language Archives Community
P-O-S	Part of Speech
Praat	A fine grained audio analysis tool
S.D.	Standard Deviation
SAMMIE	Saarbrücken Multimodal MP3 Player Interaction Experiment
SeU	Survey of English Usage
SGML	Standard Generalised Mark-up Language
SK-P	SmartKom Multimodal Corpus
SVC	SmartWeb Video Corpus
SyncTool	A tool for the transcription and alignment of audio and video
TEI	Text Encoding Initiative
The Observer	Software for coding and analysing observational data
Tractor	A digital tool for the transcription of audio and/or video records
Transana	Qualitative analysis software for video and audio data
Transtool	A digital tool for the transcription of audio and/or video records
TraSA	A digital tool for the transcription of audio and/or video records
TRP	Turn Relevance Place
VACE	Video Analysis and Content Exploitation
XML	Extensible Markup Language

Chapter 1: Introduction

1.1. General overview

Current large-scale multi-million word linguistic corpora (referred to as '3rd generation' corpora in this thesis, see Chapter 2) provide the user with an invaluable resource for generating accurate and objective analyses and inferences of the 'actual patterns of [language] use' (Biber et al., 1998: 4). They provide the apparatus for investigating patterns in the frequency of occurrence, co-occurrence (collocation) and the semantic, grammatical and prosodic associations of lexical items across large records of real-life discourse. These enquiries are difficult to undertake manually.

Consequently 'corpus analysis can be a good tool for filling us in on "the big picture" of language (Conrad, 2002: 77), providing users both with sufficient data for exploring specific linguistic enquiries, the corpora, and with the method of doing so; the Corpus Linguistic approach (see Stubbs, 1996: 41, CL hereafter).

However, current corpora have a fundamental deficiency, owing to the fact that spoken corpora are effectively mono-modal, presenting data in the same physical medium; text-based records. This is problematic because although 'we speak with our vocal organs….we converse with our whole body' (Abercrombie, 1963: 55) thus, by presenting the user with mere textual records, current corpora fail to provide adequate means for exploring communication *beyond the text*. They are inadequate in facilitating a more comprehensive investigation of not only spoken, but also non-verbal elements of language in specific contexts of communication.

It is, therefore, appropriate to propose a new, '4th generation' of corpora (see Chapter 2) and appropriate CL software to fill this void, accommodating a more *multi-modal* perspective of discourse (MM hereafter). A 4th generation MM corpus is comprised of video, audio and textual records of interaction (and associated metadata information) extracted from recordings of naturally occurring conversational episodes which are *streamed* in an easy-to-use interface; the MM corpus tool-bench. A *mode* of data, in this sense, is crudely defined as the physical format in which a particular phenomenon is presented and observed; thus, here *multi-modality* is the culmination of these integrated and aligned data streams. At this point it should be acknowledged that the literature suggests that this definition of multimodality is not without contention, however for ease of reference and to maintain consistency it *is* utilised throughout this thesis (this matter is discussed in more detail in Chapter 2).

At present, few large-scale, publicly available 4th generation corpora exist (see Chapter 2 for more information), furthermore there exists no widely established methodological approach for interrogating such datasets. To address this limitation, this thesis presents a bottom-up study of MM corpora, investigating the various issues involved in both the physical *construction* of such corpora as well as providing a worked example of an approach to the *analysis* of specific linguistic phenomena across the multiple streams of data in a MM corpus.

In terms of constructing MM corpora, this thesis examines a range of technological procedures for the collection, storage, annotation, mark-up and coding of data. Through this examination, the thesis outlines a basic way of

integrating annotations of different aspects of communicative events, in order to meet the needs of the end-user, i.e. the corpus linguistic researcher (analyst).

In addition, a corpus-based methodology for the actual investigation of these MM corpora is outlined, demonstrating how new software and methodologies can be utilised to enhance the description and understanding of language and gesture-in-talk. This provides a backdrop to the principal concern of this thesis; the investigation of patterns of the (co)occurrence of spoken backchannels and backchanneling head nods in talk (i.e.HeadTalk).

Using this MM CL approach, the thesis presents a detailed descriptive account of the relationships between the (co)occurrence of spoken and non-verbal backchanneling elements in discourse (see below for definitions). This is undertaken by testing 10 different premises, which act as research questions that were drawn up in Chapter 2, using findings from past research, together with results from the case study analysis (Chapter 4). These premises thus provide the foundations for extending what we already know about backchannels, knowledge that has previously been evidenced by text-based corpus analyses of spoken forms of this phenomenon (and related research).

Using the findings from these enquiries, that is the 'testing' of the premises, the thesis constructs a profile of the discursive roles and functions of spoken and non-verbal backchannels. This profile supplies the outlines of a detailed coding scheme/ matrix, for encoding instances of these phenomena, as well as for mapping the relationship between them. This then provides the

foundations for identifying, defining and investigating backchanneling behaviours in future research of this nature.

The thesis presents research relating to developments made during the 3-year ESRC funded DReSS (Understanding New Digital Records for eSocial Science) project based at the University of Nottingham. This interdisciplinary project involved researchers in the area of applied linguistics working in collaboration with computer scientists and psychologists, in order to develop new interfaces and tools for presenting and interrogating large quantities of MM data[1]. The thesis extends the focus of the DReSS project by providing a more detailed user-based perspective of the requirements for, and practical applications of, MM corpora (for the linguistic community), contextualising these issues through the detailed investigation of a specific linguistic enquiry.

1.2. Research questions, aims and objectives

This study therefore tackles two key issues. The first of these concerns the following:

1) *Developing the next generation of linguistic corpora*: What are the key technical, practical and ethical issues and challenges faced in the design and construction (i.e. the 'development') of MM corpora, and how can these best be approached?

In order to address this, the study concentrates on exploring the issues raised by procedures of *standardisation* in current corpora, outlining requirements for

---

[1] For more information, results and publications from DReSS, please refer to the main project website: http://web.mac.com/andy.crabtree/NCeSS_Digital_Records_Node/Welcome.html

4

the collection, transcription and codification of new data sets for MM corpora. It also addresses the need for *functionality* within MM corpora, identifying a range of technical, aesthetic and ethical problems faced when physically presenting multiple streams of audio and video data within an easy-to-use, integrated corpus tool. Drawing on these discussions, certain methodological guidelines are presented, detailing the most effective ways of addressing the issues and challenges faced in MM corpus construction, guidelines which will prove to be an invaluable source of reference in the future.

The second issue concerns the actual *usability* of MM corpora; how such corpora can be exploited to facilitate explorations of linguistic phenomena *beyond the text*. It provides an examination of the frequency of use, and relationship between, spoken and non-verbal backchannel behaviour in discourse, in order to address the second research question (which is the principal linguistic concern of this thesis):

2) Using MM Corpora: What are the roles, forms and functions of non-verbal and spoken backchanneling behaviour in real-life discourse, and what is the relationship between them?

Backchannels are described as short spoken or non-verbal response tokens that are used by a listener as a 'way of indicating….positive attention to the speaker' (Coates, 1986: 99) without attempting to take the turn in talk (see Chapter 2 for further details). Whilst ample research into spoken forms of backchannels exists, the majority of this fails to fully investigate the relevance, form and linguistic function of non-verbal backchannels in discourse.

Given this paucity, backchanneling head nods were chosen as the focus of this study because although they are to a certain extent salient and fundamental elements of spoken face-to-face interaction, owing to the sheer frequency with which nods are used (see Chapter 4), little is known about how this phenomenon actually behaves in discourse. Moreover, there is scant information regarding the interoperability of spoken and forms of non-verbal backchannels; observing how and when spoken and non-verbal forms are used alone or, in synchronicity in talk, and in determining the type of effect this has on the associated meaning of such sequences of behaviour. The second focus of the thesis is, therefore, designed to overcome these deficiencies.

The second question is addressed through assessing the following characteristics of spoken and non-verbal backchanneling behaviours:

- **Frequency**: The total number of words spoken by a participant and the total number of spoken backchannels and/or backchanneling nods they use, over the course of a conversation.
- **Type**: The patterns/ characteristics of spoken and non-verbal backchannel use, focusing on the frequency of use of:
    - Particular *forms* and discourse *functions* of spoken backchannels, and specific forms adopting a given function.
    - Individual backchanneling nod types (based on the movement structure of the nods).
- Patterns of **co-occurrence**: (i.e. the simultaneous use of spoken and non-verbal backchannels), questioning:

- Which nod types are most frequently used with spoken backchannels of a specific form and or/ adopting a specific discourse function?

1.3. Study design

This thesis is based on refining and redefining CL approaches and methodologies for MM corpora. A CL approach is an empirically-based methodological approach to the analysis of language. The central premise of CL methodology is the utilisation of large quantities of naturally occurring language-in-use as 'data', which is stored electronically as 'corpora', in order to investigate a wide range of different linguistic enquiries (see Firth, 1957; Halliday, 1978 and Sinclair, 1996). As previously stated, current CL methodologies generally deal with 'data' that comprises text-based records of language, thus are limited in providing the means for investigating MM corpora.

In order to revise current CL approaches, a critique of a range of state-of-the-art technological and theoretical methods and methodologies that enable the extrapolation and exploration of gesture-in-talk is provided. These have been taken from a variety of different disciplines of academic research (including psychology, computer science and sociology), and are adapted to meet the specific needs of the linguist, as outlined in Chapter 3. This examination provides not only the guidelines for a new approach for MM corpus development, addressing question 1, but also lays the foundations for a revised CL approach for MM data analysis, which is addressed as part of the second thesis question.

In order to examine this second question in greater detail, the thesis provides a quantitative, corpus-based numerical assessment of backchannel phenomena in naturally occurring discourse. Two different datasets are drawn upon, comprising dyadic academic supervision sessions (between supervisors and supervisees). These consist of a five-hour sample of MM video data, around 56,000 transcribed words, which is complemented by a ten-minute case study sample of the data, around 2000 transcribed words.

The case study acts as a basis for establishing a framework for examining patterns of backchanneling phenomena. It functions to determine the *how* of identifying, marking-up and comparing spoken and non-verbal backchannels, and begins to specify *what* sort of discursive and semantic functions these phenomena adopt to help generate meaning in discourse.

The approaches used in the case study are subsequently extended in the latter part of the thesis, Chapter 6, for the analysis of a five-hour, 56,000-word corpus of MM data. This corpus contains 6 complete supervision episodes, each ranging from 30 to 60 minutes in length. The characteristics of behaviours used are compared, firstly, for each speaker in each video (in order to investigate the possibility of backchanneling as an idiosyncratic activity), then across the two speakers in each dyad, before the entire dataset is analysed and more in-depth comparisons of patterns and results are undertaken.

To provide a systematic analysis of these behaviours, 10 key premises are investigated in this main study. The first 5 of these are based on a culmination of findings made from a range of different, relevant, studies of spoken backchanneling behaviour and/or gesture-in-talk from current literature.

Chapter 2 contextualises each of these premises as part of the extensive literature review. The remaining 5, which focus mainly on non-verbal backchanneling, are based on findings derived directly from the case study analysis (see Chapter 4 for further details). The 10 premises are as follows:[2]

1- 'Backchanneling occurs more or less constantly during conversations in all languages and settings' (Rost, 2002: 52, also Oreström, 1983; Gardner, 1998).

2- If one speaker dominates the conversation significantly then the other will backchannel more.

3- The simple backchanneling form *mmm* is most frequently used, based on results provided by Oreström, 1983; Gardner, 1997a, 1998 and O'Keeffe and Adolphs, 2008. The simple backchannels *yeah, okay* and *right* will also be fairly prevalent in talk, although they function in different ways to *mmm* (see O'Keeffe and Adolphs, 2008; Gardner, 1997b).

4- The continuing (CON) function is most commonly adopted by spoken backchannels, as supported by Oreström, 1983 and O'Keeffe and Adolphs, 2008.

5- Complex forms of backchannels commonly function in a more affective, relational way, than simple forms, such as *mmm* and *yeah.* These latter forms instead often function as continuer (CON) tokens, which are often more semantically empty; providing the 'most minimal' feedback (O'Keeffe and Adolphs, 2008).

---

[2] For definitions of the specific movement *structure* (described as *types*) of head nod behaviour, and the lexical *forms* and discursive *functions* of spoken and non-verbal backchannels, please see Chapters 2 and 3.

6- Backchanneling head nods are used at the same rate or more *frequently* than spoken backchannels since they are even more minimal and non-evasive than spoken forms, imposing even less of a challenge to the floor.

7- The most common *types* of head nods used in discourse are of a short duration, i.e. types **A** and **C** or less intense, multiple, type **B** nods. Types **D** and **E** are less frequently used.

8- Nods are used more frequently with concurrent spoken backchannels than alone. Similarly, spoken backchannels are used more frequently with concurrent nods than alone.

9- Spoken backchannels that are used as IR and ER tokens are more likely to co-occur with complex forms of backchanneling nods that vary with intensity, i.e. types **B**, **D** and **E,** whereas backchannels that exist on the opposite end of the 'functional cline' will co-occur with shorter, more simple, type **A** and **C** nods.

10- Spoken and non-verbal backchannels are often used collaboratively in talk, and are shown to cluster and operate in context: within and across turn boundaries.

In order to evaluate these premises, the following 5-stage approach to analysis is undertaken:

Stage 1: Specific non-verbal/ spoken behaviours are identified as backchannels.

Stage 2: The type of nod, linguistic form and discursive function of the non-verbal and spoken backchannels, respectively, are classified.

Stage 3: The frequency of each backchannel form, type and function are noted for every speaker in each supervision.

Stage 4: Instances of spoken and non-verbal backchannel co-occurrence are marked and frequencies of their associated forms and discursive functions, where known, are again noted for each speaker and across each supervision video.

Stage 5: Frequencies of spoken and non-verbal backchannels (co)occurrence from each speaker and/or video are compared and observations regarding interesting patterns made.

This study effectively adopts a mixed-method approach, combining quantitative explorations of frequencies and patterns of co-occurrence, with more a detailed qualitative assessment, a detailed discourse analytical linguistic commentary on the relevance of the results and patterns seen.

The quantitative analyses (stages 1-4) are centred on providing raw frequency counts of the occasions where features of interest are used, as well as raw percentage comparisons. This provides a simple but sufficient means of illustrating whether, for example, a pattern of behaviour seen for one speaker or dyad of speakers is similar to that shown by other speakers or across all of the supervision videos. The relative pros and cons of using this method, and alterative statistical tests, are discussed further in Chapter 6.

Thus results from such comparisons provide an adequate 'point-of-entry' into the analysis of this data.

Whilst the case study data was extracted from one conversational episode (between a young female supervisee and an older male supervisor, who are meeting for the first time in a supervision capacity at the start of an MA dissertation), the extended data-driven study includes a range of different participants, discussing a range of different topics, all of whom are at different stages in their academic careers. Some participants are meeting for the first time, whereas others are meeting during their second or final year of study. The dataset includes participants of various different ages, with examples of interactions between male-male, female-female as well as male-female dyads. This provides a larger cross section of participants than the case study and, potentially, a wider range of factors influencing the *frequency*, *type* and patterns of *co-occurrence* of spoken and non-verbal backchannels in the data.

This variation helps to provide a more socio-cultural perspective to the analysis, allowing for the examination of the potential effect that, for example, professional status has on the use of backchanneling behaviour in conversational episodes. The impacts of these socio-cultural factors are discussed as part of the qualitative assessment of the results, undertaken as part of stage 5 of the analytical approach, see Chapter 6 for further details.

1.4. Thesis overview

Chapter 2 commences with a review of past research relevant to the current study, in order to provide a theoretical and methodological backdrop to the study.

In Chapter 3, the first aim of the thesis, regarding MM corpus *development,* is examined in further detail. The standards that are used in current 3rd generation corpus design and construction are reviewed, and the necessity for adapting and extending these for the next generation of corpora is underlined. The key theoretical, ethical, practical and technological challenges faced in this redevelopment are discussed in this chapter.

The corpus development methodological focus of this chapter is complemented in Chapter 4 by a functional assessment of the issues faced in actual *usability* of MM corpora (addressing the second aim of the thesis). This chapter introduces a 3-step methodological approach for the manual definition and analysis of backchanneling behaviours, which is used to conduct a basic case study analysis of a ten-minute excerpt of MM corpus data.

Building on this initial line of enquiry, Chapter 5 questions whether the labour intensive manual methods of gesture definition and analysis used in Chapter 4 can be further enhanced to enable the analysis of large-scale MM corpora. It explores the potential for *automating the processes* of backchanneling head nod detection, definition and codification as a means of allowing the analyst to search and manipulate sizeable video datasets quickly (i.e. >100,000 words). The automated method in question is a video tracking device, the HeadTalk tracker, built by computer vision (CV) experts at the University of Nottingham (see links on the DReSS publications page for further information). The chapter questions the practical efficiently and reliability of this tracker. The chapter determines that, at present, the more analyst-led manual approach, as used in Chapter 4, is deemed to be a more accurate means of analysis.

Chapter 6 then provides a systematic investigation of the characteristics of backchannel use (according to the variables identified in section 3, above), as evidenced by the analysis of a five-hour MM corpus of dyadic supervision data. The chapter focuses on testing, supporting or refuting the claims made in the 10 key premises about backchanneling listed above. These results are complemented by a detailed linguistic commentary, in Chapter 7, profiling the key attributes of the (co)occurrence of spoken and non-verbal backchanneling phenomena in real-life interaction. This provides the foundations for a coding matrix for MM backchanneling phenomena, which is offered here, which presents guidelines for categorising and interrogating such behaviours in corpus linguistic research.

Finally, Chapter 8 provides a conclusion to the thesis, drawing all the discussions to a close. The main aims and objectives are revisited and ways in which these have been met are clearly outlined. Furthermore, the strengths and limitations of the study are explored and how these limitations might be overcome in future studies is considered.

Chapter 2: Literature Review

2.1. Introduction

This chapter lays the foundations for the exploration of the spoken and non-verbal backchannels in real-life discourse. It draws specifically on discussions of communicative *feedback* and overviews the wealth of linguistic research that explores a particular form of feedback, spoken backchanneling phenomena, in detail. Current linguistic research paradigms which can be utilised to examine the forms, roles and discourse functions of spoken backchannels are discussed in order to place the concerns of the current study in the context of previous research in the field, thus providing the basis for the investigation of research question 2.

The chapter then proceeds to highlight the minimal nature of the amount of comparable linguistic research paradigms which have been designed to enable the exploration of non-verbal backchannel behaviours. The history of research into gesture-in-talk from a variety of different disciplines is broadly examined, drawing specifically on studies of head nod behaviour and backchanneling head nods. The chapter outlines the problems faced in the definition and categorisation of the 'non-verbal' in communication, emphasising the ways in which 4th generation corpora and associated MM CL methods can perhaps help to reduce these problems by providing a platform for the systematic analysis of these behaviours.

2.2. The history of the corpus linguistic approach

2.2.1. Pre-electronic linguistic enquiry

Although the term Corpus Linguistics is relatively new, emerging around 1955, McEnery and Wilson indicate that 'the methodological ideas, motivations and practices involved in Corpus Linguistics in fact have a long history in linguistics' (1996: 1). Indeed there are many examples of early empirical studies which explored patterns of actual language-in-use. These involved scholars working with hand-written, purpose-built, collections of texts (corpora) which took an enormous amount of time and effort to design, build and analyse. Corpus studies of this nature included those examining the lexicographical and grammatical properties of language (Käding, 1879 and Boas, 1940), studies of language acquisition (see Ingram, 1978), learning and teaching (see Palmer, 1933; Fries and Traver, 1940 and Bongers, 1947) and biblical studies. Corpora in this 'pre-electronic' phase were relatively small in size, and as they were manually constructed and analysed, they were often prone to error. Yet despite this, these types of studies found support in those who believed that 'the analysis of discourse is, necessarily, the analysis of language of use' (Brown and Yule, 1983: 1).

However, as a result of these key deficiencies there were many critics of early CL methodologies (see McEnery and Wilson, 1996: 4). Perhaps the most prominent was Noam Chomsky (1965) who maintained that corpora of such small sizes were inherently unaccountable; potentially providing misleading and ungeneralisable observations of language. Consequently, he claimed that quantitative, data-driven, empirical investigations of language undertaken using corpus-based methodologies were essentially 'skewed'. He

proposed that more rationalistic approaches, involving the qualitative assessment of introspective data, exist as more reliable techniques for linguistic analyses. This supported Bourke's view that 'the criterion of truth is not sensory but intellectual and deductive' (1962: 263).

As a result, there was a general movement from empiricism to rationalism in linguistic research at this time. This was described as the 'time of discontinuity', where Corpus Linguistics generally fell out of favour with linguists, although research using CL based methods did not halt completely (McEnery and Wilson, 1996: 4).

2.2.2. 2nd and 3rd generation Corpus Linguistics

Since these early days, empirical language analyses have witnessed a resurgence in popularity, one which can primarily be attributed to the advent of computers.

The 2nd generation of corpora, early computerised corpora which embraced the 'digital age' in its early stages (until around the 1980s), revolutionised the potential of CL enquiry by enabling linguists to systematically create digital records of corpus data on-screen. They also enabled digital searches of the corpora, rather than the researcher having to trawl through numerous pages of hand-written accounts when analysing data, thus dramatically reducing the time and accuracy with which these enquiries were undertaken. The most renowned of 2nd generation corpora are the Brown corpus, built in 1963, and the Lancaster-Oslo/ Bergen corpus (LOB), built in 1975, a computerised version of components taken from the SeU (Survey of English Usage).

The ever-increasing sophistication of computers, over the past thirty years in particular, has now provided linguists with the ability to compile and 'handle huge amounts of [corpus] data' (Kennedy, 1998: 5). Large computerised corpora of this nature are defined as 3rd generation corpora within this thesis and include major corpora to date, such as the British National Corpus (BNC-100 million words of written and transcribed spoken discourse), the 524 million-word Bank of English (BoE) and the Cambridge and Nottingham Corpus of Discourse in English (CANCODE[3], a 5 million word corpus of transcribed spoken data).

Furthermore, the development of digital concordancing software, such as Wordsmith Tools (Scott, 1999), has also enabled researchers to 'calculate frequencies, analyse collocates and often calculate statistical measures of the strength of word associations' (Conrad, 2002: 77) with 'incredible speed, total accountability, accurate replicability [and] statistical reliability' (Kennedy, 1998: 5). This has enabled the following types of enquiries about language to be undertaken with relative ease (Conrad, 2002: 77-83):

1. Investigating characteristics associated with the use of a language feature.
2. Examining the realisations of a particular function of language.
3. Characterising a variety of language.
4. Mapping the occurrence of a language feature through a text.

---

[3] CANCODE stands for Cambridge and Nottingham Corpus of Discourse in English, a 5 million word corpus of spoken English taken from different contexts across the British Isles. CANCODE was built in collaboration by The University of Nottingham and Cambridge University Press (with whom sole copyright resides).

The various developments in modern Corpus Linguistic methodology have provided a research landscape that contemporary CL enthusiasts believe stands in strong opposition to Chomsky's alternative, rationalistic approach to language analysis. They cite the following as particular advantages of using CL methodologies (based on McEnery and Wilson, 1996: 8, also see McCarthy, 2001: 125 and Meyer, 2002: 5):

- Introspective data is artificial, whereas corpora are natural, unmonitored sources of data.
- Frequencies of word / phrase / grammatical construction use cannot be discovered without the use of a corpus. In order to obtain frequency information, CL techniques are the only option, as human beings have only the vaguest notion of the frequency of lexical units and thus need to draw on the naturally occurring data to make accurate statements about word frequency.
- The process of introspection may not be at all systematic and is definitely less systematic than a corpus approach.

As a result, while 'artificial data can have a place in modern corpus linguistics…. it should be used with naturally occurring data which can act as a control, a yard stick if you will' (McEnery and Wilson, 1996: 16).

Corpus-driven approaches are now used in a wide range of linguistic disciplines, including semantics, pragmatics, stylistics, language learning and pedagogy, and can be adapted beyond the field of linguistics, for example, in psychological and sociological motivated investigations of language-in-use.

2.2.3. 4th generation Corpus Linguistics

Despite these strengths, as outlined earlier in the introduction, current spoken corpora are limited as they only have the provision for presenting data in a single format; that is text, in the form of transcripts of interactions (see Knight et al., 2006). They provide little opportunity for exploring non-verbal, gestural aspects of discourse because 'the reflexivity of gesture, movement and setting is difficult to express in a transcript' (Saferstein, 2004: 213). Thus, these text-based accounts of interaction only allow for a partial description of discourse delivered through corpus analyses (Wilcox, 2004: 525). 4th generation MM corpora aim to overcome this partiality.

As Lund notes (2007: 289-290):

The term multimodality encompasses a wide variety of phenomena in the literature, including emotions and attitudes conveyed through prosody, applause, laugher or silence in answer to a question, body movements, object manipulations and proxemics, layout and posture……in a different vein, the term multimodal is also often used to signify the medium in which a particular message can be expressed, for example text and graphics.

Therefore, when discussing MM research and MM corpora, essentially, we are looking not only towards the '*abstract*' elements in discourse; the processes of 'meaning making' (i.e. bodily movement and speech, see Kress and van Leeuwen, 2001), but also the '*media*', the physical mode(s) in which

these *abstract* elements are conveyed. Thus, as Lund notes, since 'the mode of gesture is carried out by the media of movements of the body' (2008: 290, paraphrased from Kress and van Leeuwen, 2001), it seems logical to define the multi-modal as a culmination of these senses of the *abstract* and the *media*.

So while MM behaviours (in interaction) are involved in the processes of meaning generation, the MM *corpus* is the physical repository, the database, within which records of these behaviours are presented, through the culmination of multiple forms of *media,* i.e. different *modes* of representation. Thus, the 'multi-modal corpus' is defined as 'an annotated collection of coordinated content on communication channels including speech, gaze, hand gesture and body language, and is generally based on recorded human behaviour' (Foster and Oberlander, 2007: 307-308). The integration of textual, audio and video records of communicative events in MM corpora provides a platform for the exploration of a range of lexical, prosodic and gestural features of conversation (the *abstract* features, see Kress and van Leeuwen, 2001), and for investigations of the ways in which these features interact in real, everyday speech.

2.2.4. Current multi-modal corpora

4th generation MM corpora are still very much 'under development' (referred to as *developing* hereafter), and as yet no ready-to-use large corpus of this nature is commercially available. This is owing to a variety of factors, but, principally, due to 'privacy and copyright restrictions' (van Son et al., 2008: 1). Those that have been built are often designed to fulfil particular aims of a

research project (see Knight, 2006) and have limited functionality beyond such a project (refer to Figure 2.1 for an index of these).

Current MM corpora also tend to only feature a small number of participants and/or focus on a specific discourse context, providing little utility for describing language use beyond this context. Examples of such include the Fruits Cart Corpus and MIBL Corpus, as detailed in Figure 2.1.

In addition such MM corpora, including the IFADV corpus and Göteborg Spoken Language Corpus for example, are also relatively limited in size, especially compared to the multi-million-word 3rd generation corpora already in existence. Even though the largest MM corpus documented in Figure 2.1, the AMI corpus (see Ashby et al., 2005) comprises an impressive 100 hours of video, the majority of this data exists solely as video records. In other words all of the videos have yet to be transcribed, thus the actual size of this corpus, as a functional MM (i.e. text and video based) tool is not especially large.

However, although one limitation of current MM corpora, related to size, has been postulated here, it is in fact not necessarily valid to gauge size in relation to pre-existing textual corpora. As it is important to note that 'what is meant by large corpora is however quite a relative notion' in linguistic research (Blache et al., 2008: 110). 'In some linguistic fields such as syntax for instance, corpora of several million words are used, whereas in prosody where most of the annotations are made manually, a few hours of speech are considered as a large corpus' (Blache et al., 2008: 110). Therefore, the appropriateness of size can only really be determined in light of the specific researcher's requirements. Thus, caution should be exercised when qualifying size as a strength or key shortcoming of a MM corpus.

Name and Language Origin	Size, Composition and Additional Information	Reference(s)
AMI Meeting Corpus (non-native English)	100 hours of recordings taken from 3 different meeting rooms. This corpus was created for the use 'of a consortium that is developing meeting browsing technology'.	Ashby et al., 2005
CID (Corpus of Interactional Data)	8 hours of dyadic conversations, comprised of 2 participants sat in close proximity of one another, each wearing a microphone headset. Participants were encouraged to chat informally, so with no directions on how to structure the talk- promoting spontaneous discourse.	Bertrand et al., 2006; Blache et al., 2008
Czech Audio-Visual Speech Corpus / Corpus for Recognition with Impaired Conditions	Developed to test and train the 'Czech audio-visual speech recognition system' (automatic speech recognition). The first corpus features 25 hours of audio-visual records, from 65 speakers, the second has 20 hours of data across 50 speakers. In both, each speaker was instructed to read 200 sentences each, in laboratory conditions (50 common, 150 specific to the speaker).	Železný et al., 2006; Trojanová et al., 2008
Fruits Cart Corpus	104 videos of 13 participants, 4-8 minutes each = approx 4000 utterances in total. Comprised of task-orientated dialogues in an academic setting. Designed to explore language comprehension, now used to analyse language production (NLP research).	Aist et al., 2006
Göteborg Spoken Language Corpus	Small components of this 1.2 million word spoken language corpus have been aligned with video records. Conversation is taken from various different social contexts with a range of different speakers talking 'spontaneously'.	Allwood et al., 2000
IFADV Corpus	A free dialog video corpus composed of face-to-face interaction between close friends/ colleagues. This corpus is comprised of twenty 15 minute conversations (5 hours in total).	Van Son et al., 2008
MIBL Corpus	Comprised of human-to-human instruction dialogues, with one participant teaching a card game to the other (similar to map task activities, see the Map Task Corpus, Anderson et al., 1991). This corpus links speech to movement on the screens and is used to train service robots ('corpus based robotics').	Wolf and Bugmann, 2006
Mission Survival Corpus 1 (MSC 1)	A meeting corpus which includes a range of short meetings, with up to 6 participants in each. The topics and tasks covered in the meetings are controlled but not scripted.	Mana et al., 2007
MM4 Audio-Visual Corpus	Features 29 short meetings between 4 people filmed in controlled, experimental conditions. The majority of the meetings were scripted and cover specific, predetermined, topics and tasks.	McCowan et al., 2003
NIST Meeting Room Phase II Corpus	Part of the NIST MDCL (Meeting Data Collection Laboratory). This corpus contains 15 hours of recordings from 19 meetings; comprised of both scenario-driven meetings and 'real' meetings.	Garofolo et al., 2004
NMMC (Nottingham Multi-Modal Corpus)	250,000 words, 50% single speaker lectures, 50% dyadic academic supervisions. Sessions were video and audio recorded, transcribed and aligned using DRS (the Digital Replay System).	Knight et al., 2009
SK-P 2.0- SmartKom Multimodal Corpus	96 different single 'users' were recorded across 172 sessions, each recorded in public spaces such as at the cinema or in a restaurant. Sessions were video and audio recorded. HCI.	Schiel et al., 2002
SmartWeb Video Corpus (SVC)	99 recordings of human-human-machine dialogue, i.e. 1 speaker (recorded) interacting with a human person and a dialogue system (i.e. the main participant is using a Smartphone, which records their face and they are talking to the other participant).	Schiel and Mögele, 2008
VACE Multimodal Meeting Corpus	Comprised of meeting room-based 'planning sessions'. Spontaneous talk in controlled environments (participants given specific tasks to fulfil). 5 participants present in each scenario, across 5 scenarios.	Chen et al., 2005

Figure 2.1: An index of multi-modal corpora.

AMI, as with the MM4 Audio-Visual Corpus, MSC1, the VACE Multimodal Meeting Corpus and the NIST Meeting Room Phase II Corpus all feature records of interaction extracted from one specific discourse context, a professional meeting room. In these meeting-based corpora, the primary motivation behind the associated research (and corpus construction) is to enable the development and integration of technologies for displaying and researching meeting room activity. In some of these corpora, the content is scripted or pre-planned to a certain extent and/or the conditions in which the recordings take place are controlled and experimental, with participants being told specifically where to sit, etc.

So, the AMI, despite its commendable size, together with the other meeting corpora seen here, are heavily 'specialist', and thus are limited in their usefulness for general CL research because they are so contextually and compositionally specific. 3rd generation specialised corpora are similarly commonplace, such the MICASE corpus[4] of academic discourse, and the Wolverhampton Business English Corpus[5]. These corpora are not necessarily appropriate for addressing research questions that focus on the more interpersonal aspects of communication, beyond this formal, professional contextual domain. This is because the meeting room environment is generally regarded as not being particularly conducive to the frequent occurrence of more informal, interpersonal language and/or behaviours.

---

[4] MICASE, the Michigan Corpus of Academic English, is a 1.7 million word corpus of transcribed interactions recorded at the University of Michigan. For more information see: http://lw.lsa.umich.edu/eli/micase/index.htm

[5] The Wolverhampton Business English Corpus is comprised of 10 million words of written English from the business domain. These texts were collected between the years 1999 and 2000. For more information see: http://www.elda.org/catalogue/en/text/W0028.html

It is relevant to note that it is not merely the specialist business corpora, as featured in Figure 2.1, that are perhaps inappropriate for analysing patterns of spontaneous, innate language use. Both the Czech Audio-Visual Speech Corpus and the Czech Audio-Visual Speech Corpus for Recognition with Impaired Conditions contain purely scripted speech, with participants reading lines of text, rather than featuring naturally occurring interaction. Similarly, the SK-P was filmed in 'wizard-of-oz' settings (see Schiel et al., 2002 for further details), with single participants instructed to carry out particular tasks (interacting with a computer screen, HCI- Human-Computer-Interaction), rather than engaging in naturalistic multi-party conversation.

While the SK-P is useful for training and testing simulated dialogues systems, and for other research into specific forms of HCI (Schiel et al., 2002, also see the SAMMIE[6] corpus, Kruijft-Korbayova et al., 2006), it has little use beyond this.

As identified in Chapter 3, the NMMC aims to provide the linguist with more naturalistic forms of MM data, away from the scripted content and experimental conditions used in many of the corpora mentioned above. However, it should be noted that although there are plans to record data from a 'range of contexts' in the NMMC, this aim has yet to be realised. Currently, the data contained within the NMMC is somewhat specialised and context-specific, in that it only features recordings from academic supervisions. However since it is the only corpus freely available for the purpose of this thesis research, it was the one selected for use here.

---

[6] SAMMIE stands for the Saarbrücken Multimodal MP3 Player Interaction Experiment. For more information see: http://www.dfki.de/lt/publication_show.php?id=4041

The most 'naturalistic' of these corpora are the SVC and CID. The SVC is described as containing records of 'spontaneous' communication between two participants. However, participants are not strictly involved in face-to-face dialogue. They are instead using Smartphones as a medium for interacting, which simultaneously record the voices and facial images of participants as they talk to each other. Consequently, this corpus can also be described as specialist, as although speakers can see each other, the data comprises human-human-machine interaction, rather than merely human-human. Thus, the transferability of this data beyond this particular context of recording (human-human-machine video phone based dialogue) is again restricted.

In contrast the CID is comprised of real-life interaction between two people sitting next to each other, who are encouraged to discuss any topic or issue they wish. This means that the CID is a more general corpus than those discussed so far, as it seeks to provide conversational data in German (making this particular corpus unsuitable for use here, since this study is focused on English language-in-use) which is as naturalistic and context free as possible. However, the conditions in which these recordings took place are to a certain extent experimental, with participants sitting in a laboratory and wearing headset microphones. Although obviously it can be widely debated whether or not the headsets actually compromise the 'naturalness' or authenticity of the data contained within this corpus.

Overall, the various 'shortcomings' attributed to these current MM corpora, from the AMI to the VACE corpus, can perhaps be attributed to the fact that current 4th generation MM corpus research projects, in general, tend to concentrate on only one of the following concerns (taken from Gu, 2006: 132):

A: Multi-modal and multimedia studies of discourse

B: Speech engineering and corpus annotation

Specific examples of the former (A) include work by Kress and Van Leeuwen (1996), Martinec (1998, 2001), Scollon (1998), Krauss (2002), and Gripsrud (2002). These studies emphasise the importance of Firth and Malinowski's notion of the 'context of situation' (see Malinowski, 1923 and Firth, 1957) and seek to explore how 'different semiotic modalities make different meanings in different ways according to the different media of expression they use' (Baldry and Thibault, 2006: 4). They concentrate on actually exploring patterns of behaviour, the *abstract* elements in discourse (refer back to section 2.2.3, Kress and van Leeuwen, 2001), in different discursive contexts through different technological *media*. Studies of this nature, therefore, use MM corpora as a means to an end, as a method for actively exploring a particular research question or aim.

Conversely, current MM corpora and corpus projects that focus heavily on the second concern (B) tend to concentrate on actually developing software and hardware tools, the *media* (Kress and van Leeuwen, 2001), for exploring language behaviours in given contexts of communication, that is facilitating the kind of research conducted by the former type of project. These studies generally support the physical construction of a corpus in some sort of capacity, but one that is limited in utility beyond that immediate discourse context.

27

In other words, these studies aim to explore *how* language in *x* context can best be captured, investigating *what* technological resources can assist researchers in realising this aim. Such corpora ultimately fail to provide in-depth investigations of specific linguistic characteristics of this data. This is particularly true of the abundant amounts of meeting corpora that exist, as discussed above. Examples of other researchers working towards this latter concern include Gibbon et al. (1997), Hill (2000), Taylor et al. (1999) and Allwood et al. (2001).

Few studies concentrate on both of these key concerns in great detail. This is mainly due to the fact that different types of expertise are needed to meet the requirements posed by each of these strands of research. While the former is conventionally undertaken by those 'in the social sciences' who 'are interested in human beings' the latter is more often the concern of computational linguists and computer scientists primarily interested in 'how to improve human–computer interaction', i.e. for developing the tools for researching the former without actually putting these tools into any great use (Gu, 2006: 132).

Nevertheless, it is appropriate to note that there has in fact been a recent surge in the number of research projects that look towards combining these two concerns. These projects (an example of which is the ongoing DReSS project which has built the NMMC, the Nottingham Multi-Modal Corpus[7]) are concerned with building on the strengths of current 3rd generation corpora by constructing corpora that are large in size and sourced from a range of discourse contexts, as well as extending the utility of 3rd generation corpora

---

[7] Refer to the main DReSS website for further details of the NMMC, and further publications linked to the DReSS project:
http://web.mac.com/andy.crabtree/NCeSS_Digital_Records_Node/Welcome.html

by providing the means of investigating language beyond the text. With the ever-increasing sophistication of digital technologies, it is assumed that the landscape of MM corpora is likely to undergo many dramatic changes and developments in the next decade or so.

The fact that the current thesis has a multi-dimensional research focus, i.e. combining the construction *and* use of MM corpora, means that it aims to finely align these two research concerns by identifying how the latter can help to inform the former. The research conducted, therefore, offers methodological blueprints and advice for 'best practice' (requirements of *developing* MM corpora are outlined accordingly).

The more technological concerns of corpus construction are investigated in Chapter 3, while the remainder of this chapter considers the theoretical background for exploring the linguistic characteristics and properties of human communication, providing the foundation for the investigation of backchanneling behaviour in later chapters.

2.3. The functions, forms and relevance of backchanneling

2.3.1. Communication and communicative feedback

2.3.1.1. Conceptualising communication

As a starting point for the discussion of backchannels, it is logical to examine research into this phenomenon from a top-down perspective, and contextualise backchannels within the wider linguistic landscape of *communication*. Early linguistic models conceptualised communication as a linear process (see Clark and Krych, 2004 for further discussions on this

matter), as illustrated in Shannon and Weaver's model of communication in Figure 2.2 (published in 1949).
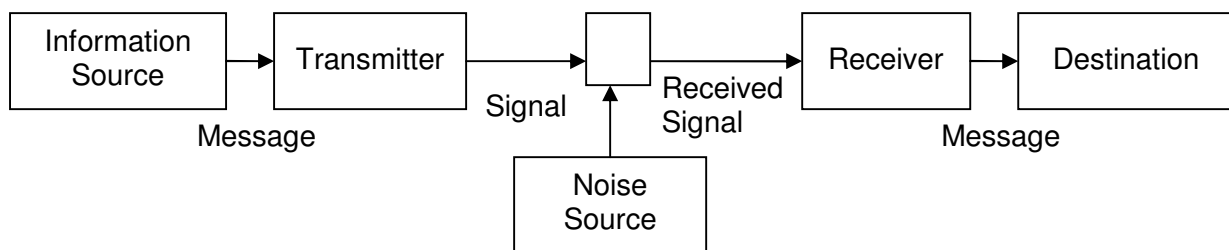
Figure 2.2: Shannon and Weaver's 'General model of communication'.

This model depicts spoken communication as containing five key elements, comprising the information source, the transmitter, the noise source, the receiver and the destination. According to this model, an 'information source' is encoded by the speaker to form a 'message' which is subsequently 'transmitted' as a 'signal', using a specific channel of communication. In this model it is a spoken channel, resulting in the 'noise source'. This signal is received by the listener (the 'receiver') and is decoded as a 'message' by the listener, thus reaching its 'destination'. Essentially, this model depicts an input (a sense or an idea) which is delivered by the speaker, and following various schematic processes, is heard by the listener who, it is presumed, successfully understands the message before decoding it (see Clark and Schaefer, 1989: 260-263).

This cumulative process is mapping a theoretical optimum, a one-to-one relationship between the starting point where the input is given, and the end point where the output message is received. In reality, in real-life communication there is not always a congruity between the input message

and the source output, and this one-to-one relationship is not necessarily always maintained. For example, a pragmatic failure may occur where listeners fail to hear or understand the message delivered by the speaker, subsequently causing problems at the encoding and/or decoding phase (see Thomas, 1983 for further details on pragmatic failure in discourse).

Similarly, the roles of the speaker and listener (recipient) in communication, and the relationships between them, are far more complex than this model suggests. Since 'speech acts are directed at real people' (Clark and Carlson, 1982: 335), real-life conversations are regarded as being more co-operative, 'highly co-ordinated activities' than the model suggests (Clark and Schaefer, 1989: 259).

Speakers are not merely poised to deliver a message, instead they actively 'monitor not just their own accounts, but those of their addressees, taking both into account as they speak' in conversation (Clark and Krych, 2004: 66). Moreover, 'addressees' are not merely passive repositories of information delivered by the speaker, but they also, 'in turn, try to keep speakers informed of their current state of understanding' (Clark and Krych, 2004: 66). This notion of the 'informed understanding', of whether and how a listener comprehends a particular message is often signalled as communicative feedback (i.e. transmitted from the 'destination', the 'receiver' back to the 'source), using not only words but also but 'non-verbal means like posture, facial expression or prosody' (Allwood et al., 2007b: 256). Instead of relying simply on *inputs* and *outputs* (as a one-to-one relationship), this additional process of feedback implies that communication is more

appropriately conceptualised as a *cyclical* process (see Patton and Giffin, 1981: 6).

As a result of the various processes of feedback, the listener 'has a crucial influence' (McGregor and White, 1990:1) on shaping interactions. This helps to provide 'strong grounds for conceptualising language [and communication] as intrinsically social' (Goodwin, 1986: 205, also see Halliday, 1978 for discussions of communication as a 'social semiotic system': an idea discussed in more detail in section 2.3.2.4), with meaning being constructed and comprehended in a variety of ways, beyond merely the choice of words spoken (i.e. delivered from 'input' to the 'output').

Feedback is seen to operate in a variety of different ways in discourse. The key functions of feedback are classified by Allwood et al. according to three basic 'behaviour attributes' (2007a: 275, these annotations exist as part of the MUMIN coding scheme- A Nordic Network for MUltiModal Interfaces). The most 'basic' attributes are the signalling of 'continuation, contact and perception' (CP) and the signalling of 'continuation, contact, perception and understanding' (CPU), where the interlocutor 'acknowledges contact' with the speaker, and in the case of CPU, demonstrates whether they understand the message or not (Allwood et al., 2007a: 275, also see Allwood et al., 1993; Cerrato, 2002; Cerrato and Skhiri, 2003 and Granström et al., 2002 for alternative coding schemes for (non)verbal feedback in communication). This can be followed by the additional attributes of 'acceptance' and 'additional emotions/ attitudes' being expressed by the listener (Allwood et al., 2007a: 275, also see Cerrato, 2004). These attributes of feedback are summarised in Figure 2.3 (taken from Allwood et al., 2007a: 276):

Behaviour Attribute	Behaviour Value
Basic	CP CPU
Acceptance	Acceptance Non-acceptance
Additional emotion/ attitude	Happy, sad, surprised, disgusted, angry, frightened, other

Figure 2.3: Feedback attributes (from Allwood et al., 2007a).

While modern models of communication acknowledge that start and end points do exist in communication, insofar as there have to be openings and closings to communication otherwise this would suggest that humans never ever stop communicating, these are complemented by, amongst other things, networks of feedback. This equates in a more dynamic and pragmatic viewpoint to a socially determined and integrated communicative process.

2.3.1.2. Contextualising backchannels

Modern models of face-to-face communication generally agree that various key 'universal' elements exist as a means of framing and structuring conversations. These are summarised by Goffman in the following list (1974), these elements can be sub-classified into various other discourses processes, as discussed below:

1. Openings

2. Turn-Taking

3. Closing

4. Backchannel Signals

5. Repair systems

The second element, turn-taking, is considered central to the management of conversation. Turn-taking is conventionally defined in predominantly lexical terms, with no consideration of non-verbal counterparts, and has been widely researched as part of the CA (Conversation Analysis) research tradition. The most comprehensive account of turn-taking is provided by Sacks et al. in 1974. Other seminal works on this phenomenon include that by Yngve (1970), Duncan (1972), Allen and Guy (1974), Goffman (1974) and Argyle (1979).

A turn is defined as 'the talk of one party bounded by the talk of others' (Goodwin, 1981: 2). During turn-taking the prospective speaker (i.e. the hearer/ listener at a given point in the conversation) is either 'nominated' by the current speaker or 'self-selected' to take the floor, the turn, from the former speaker. This marks a transition of the participant's role from listener ('recipient', see Sacks et al., 1974) to speaker (interlocutor) in the conversation. Situations where the receiver neglects to either be nominated or self-selected to 'elicit the continued speakership of the previous speaker' (Houtkoop and Mazeland, 1985: 605- based on Sacks et al., 1974) are described as marking a 'continued recipiency' role for that participant. Situations witnessing either a transition of a participants' role from listener to speaker, or a topic change, are regarded as points of 'speaker incipiency' (Jefferson, 1984).

Thus, during turn-taking 'one party talks at a time and, though speakers change, and the size of turns and ordering of turns vary; transitions are finely coordinated' (Sacks et al., 1974: 699). This is because 'the structure of the discourse is cooperative and utterances from all the participants contribute

towards its construction' (Sinclair, 2004: 104, also see Grice's maxims of cooperation, 1989 and McCarthy, 2003: 33).

In Figure 2.4 (from the NMMC) the notion of turn-taking is crudely taken, for illustration purposes, as the transition between individual 'utterance units' (Fries, 1952: 23); the chunks of talk identified after the speaker tags (**<$1>** and **<$2>**).

```
<$2> <$=> Oh well I <\$=> I'm just reading things at the moment and
      just+
<$1> Right.
<$2> +kind of vague+
<$1> So what given that the amount of stuff on metaphor is huge?
<$2> Yeah well I've been looking through some of the stuff on scientific
      metaphors and+
<$1> Uh-huh.
<$2> +er particularly how they're used for educational purposes in+
<$1> Right.
<$2> +explaining concepts erm+
<$1> Yeah.
```

Figure 2.4: An excerpt of a transcript of dyadic communication.

Given the comments and definitions discussed previously, theoretically the excerpt comprises 10 individual turns, organised systematically with each new 'speaker' following the last, so with 5 from each speaker. However, as explored in more detail below (see section 2.3.1.3), this crude alignment of utterance = turn is somewhat misleading and has been largely discredited across literature in the Discourse Analysis (DA) tradition (see, for example, Sacks et al., 1974). This is because real-life conversations also contain, amongst other things, 'backchannel signals' (refer back to the Goffman model, 1974). Backchannels are discourse phenomena that are closely related to

turn-taking, although provide an 'antithesis' to the utterance = turn dichotomy (Mott and Petrie, 1995).

The term 'backchannel' was first coined by Yngve (1970), but is also known by a variety of different terms including 'accompaniment signals' (Kendon, 1967), 'listener responses' (Dittman and Llewellyn, 1968 and Roger et al., 1988), 'assent terms' (Schegloff, 1972), 'newsmarkers' (used by Gardner, 1997a, when describing a specific type of backchannel), 'receipt tokens' (Heritage, 1984), 'hearer signals' (Bublitz, 1988), 'minimal responses' (Fellegy, 1995) and 'reactive tokens' (Clancy et al., 1996).

Yngve observed that 'when two people are engaged in conversation they generally take turns' [but] 'in fact, both the person who has the turn and his partner are simultaneously engaged in speaking and listening…. because of the existence of what I call the 'backchannel'' (Yngve, 1970: 568).

Backchannels 'help to sustain the flow of interactions' (Oreström, 1983: 24). They exist to reinforce Grice's maxim of co-operation in talk (1989), by allowing the listener to signal attention to the speaker (i.e., they are 'non-floor-holding devices', O'Keeffe and Adolphs, 2008: 74) without interrupting the flow of conversation. So, as candidly observed by Oreström, 'while a turn would imply 'I talk, you listen' a backchannel implies 'I listen, you talk' (1983: 24). Thus it can be suggested that if one speaker, engaged in dyadic conversation, is more vocal, significantly dominating the talk, then the other participant is likely to backchannel more[8].

---

[8] This finding provides the stimulus for premise 2, see Chapter 5 for details.

In addition to maintaining the conversational 'flow', backchannels also help to mark convergence and maintain relations across the speakers (an idea which was also explored by Watzlawick et al., 1967); that is, functioning both organisationally and relationally in discourse (O'Keeffe and Adolphs, 2008: 87). However, it is important to note that such backchannels 'are normally not, if ever, picked up on and commented on by the other speaker' (Oreström, 1983: 24), although turns commonly, although not always, are.

Many different *types* of backchanneling behaviour exist in conversation. These include a variety of different verbal, vocal and gestural signals, a combination of which may be used simultaneously at a specific point in talk. Duncan and Neiderehe categorise the different types in the following way (1974, see also Duncan and Fiske, 1977 for a similar categorisation scheme):

1. Readily identified, verbalised signals such as *yeah, right, mmm*

2. Sentence completions

3. Requests for clarification

4. Brief restatements

5. Head Nods and shakes

Although a wealth of linguistic research exists into the first 4 of these (for examples of such see Clark and Schaefer, 1989; Allwood et al., 1993; Drummond and Hopper, 1993a, 1993b; Fellegy, 1995 and Lenk, 1998, most which exist in the CA tradition), 'little work accounts for the [more] multi-modal character of backchannels' (Bertrand et al., 2007: 1), that is backchannels of type 5 on the above list. However, since this present section is focused

specifically on spoken backchanneling behaviour, only the first 4 types are of concern here, while head nods and shakes are discussed in section 2.4.

2.3.1.3. Backchannels Vs turns

Using the first 4 categories in this list it is possible to identify 4 instances of spoken backchannel behaviour in the transcript excerpt in Figure 2.4 (marked in **blue**). These are identified in Figure 2.5.

```
<$2>  <$=> Oh well I <\$=> I'm just reading things at the moment and
       just+
<$1>  Right.
<$2>  +kind of vague+
<$1>  So what given that the amount of stuff on metaphor is huge?
<$2>  Yeah well I've been looking through some of the stuff on scientific
       metaphors and+
<$1>  Uh-huh.
<$2>  +er particularly how they're used for educational purposes in+
<$1>  Right.
<$2>  +explaining concepts erm+
<$1>  Yeah.
```

Figure 2.5: Defining backchannels in a transcript excerpt.

In this figure *right, uh-huh, right* and *yeah* are defined as 'readily identified, verbalised [backchannel] signals' (Duncan and Neiderehe, 1974), which function to provide feedback to speaker **<$2>** without a movement to take over the floor (i.e. recipiency is maintained).

It is important to note that whilst the response *yeah* occurs twice in a turn-initial position in this excerpt, only one (the second *yeah*, marked in **blue**) of these instances is actually denoted as being a backchannel. In the second instance, *yeah* is simply used to indicate that the interlocutor is listening and

wishes the speaker to continue the conversation. In contrast, the first use of _yeah_ is used as a signal for the interlocutor to take the turn. That is, to signal the move to speakership, given that it comprises part of a full turn and is thus followed by additional talk.

Similarly, while _right_ and _uh-huh_ (marked in **blue**) are purportedly used as backchannels in this instance, they are not necessarily always examples of backchanneling behaviour when used in other situations across talk (even if used by the same speaker). As with _yeah_, they may also exist as either part, or indeed the entirety, of a turn, and indeed even is used in isolation, with no subsequent speech, it is not necessarily the case that a backchannel has been uttered. Thus, in terms of _form_, turns and backchannels can in fact both be simple, brief contributions with minimal semantic content, although not always, as discussed in section 2.3.2.1.

In terms of _location_, both turns and backchannels are turn-initial elements. Furthermore, many backchannels are also often positioned at Transition Relevance Places (TRPs, taken from Sacks et al., 1974). TRPs are where turn exchanges can, in accordance with Grice's maxims (1989), appropriately occur without being evasive and interrupting the cooperative nature of the conversation. These are points where 'the current hearer can [theoretically] take over the main channel of communication by taking a turn' (Cathcart et al., 2003: 52). If a further contribution is not made at the TRP, following the listeners' given utterance, the contribution can often be legitimately classified as a backchannel. However, with its location at the TRP position, a turn may also relevantly be initiated instead of a backchannel.

Given these similarities, it is appropriate to question how one can effectively establish whether a minimal response in talk exists as a backchannel or a turn. In reality, there is not a wholly straightforward to answer this question, as the two phenomena are not strictly 'mutually exclusive' (Allwood, 2007a: 279). Furthermore, the dynamic and elusive nature of real-life communication means that it is difficult to develop 'precise and replicable tools for labelling recipients [brief] contributions' (Sacks et al., 1974, also see Duncan and Niederehe, 1974 and Goodwin, 1981: 15), making exact specifications for turn and backchannel classification, definition and differentiation problematic.

This problem of definition is further compounded by the fact that, as with full turns, 'backchanneling occurs more or less constantly during conversations in all languages and settings' (Rost, 2002: 52)[9]. Gardner concludes that spoken forms alone, 'can occur more than a thousand times in a single hour of talk' (Gardner, 1998: 205), a rate which is supported by Oreström who suggests that 8 out of 10 spoken backchannels made in conversation are emitted within 1-15 seconds of each other (1983: 121), although this number naturally varies across speaker and context. Since turns are equally as frequent in talk, the successful definition of backchannels cannot rely on frequency information alone.

As a result, it is necessary to search for additional 'clues' to assist in the profiling of a contribution (in addition to lexical form and frequency), in order to distinguish whether it is a backchannel or turn. Working on the notion that 'you shall know a word by the company it keeps' (Firth, 1957: 11, also see Tottie,

---

[9] This finding provides the stimulus for premise 1, see Chapter 5 for details.

1991: 260), an examination of the immediate lexical co-text of the utterance, that is the exact 'point' in which a lexeme or utterance is positioned in talk, observing what occurs before and after it, as well as its wider discursive context, for instance, how the utterance is framed in relation to the wider conversational episode, can assist in this definition. Similarly, the examination of concurrent non-verbal behaviours, such as sequences of gesture and facial expressions (see Allwood et al., 2007b: 256), are critical in defining backchannels in talk, something which MM corpora aim to facilitate.

Furthermore, the status of a contribution as a spoken backchannel is, to a certain extent, dependent on the prosodic characteristics of the specific lexeme or utterance. That is, the patterns in pause phenomena of lexical elements that occur before and after the backchannel, and the general 'timing' of speakers in the conversation (see Müller, 1996; Stubbe, 1998b and Grivicic and Nilep, 2004 for examples of studies that examine prosody and intonation in relation to backchannel positions and forms).

Gardner illustrates this point with an examination of three common spoken backchannel forms (see 3.3.1. for further details), *mm hm, yeah* and *mm.* He states that the 'typical' prosodic properties of each of these forms when functioning as backchannels are as follows (1998: 216):

- *mm hm* is typically marked by a falling-rising pitch contour in speech
- *yeah* and *mm* are adopt a falling intonational contour

Although on occasion it is possible for each form to 'take a different contour' depending on their respective roles or functions in discourse, Gardner

advocates that in instances where the given utterances possess these prosodic patterns, it is likely that backchanneling is taking place (Gardner, 1998: 216).

The close analysis of prosodic characteristics can also assist in informing us of the more specific discursive function that a backchannel fulfils, since spoken backchannels are used to adopt a variety of roles in discourse, as discussed in section 2.3.2.2. Müller suggests that the backchannels which function in more supportive ways in conversation, i.e. those with a higher semantic 'content' than other backchannel forms, are 'more varied in intonation, in lexical selection and also in length' (1996: 163, cited in Kendon, 1997). Gardner illustrates this point with the suggestion that, for example, backchannels with 'a marked rise-falling tone or high pitch' (i.e. the example of *mm hm* given above) are more likely to be 'used to express encouragement or appreciation, or if low and level in tone, indifference' (2001: 13) than those with a falling contour.

When defining turns and backchannels in the excerpt seen in Figure 2.5, it was possible to examine the patterns of pause phenomena around these, by simply replaying this time-aligned extract with the corresponding audio recording of the supervision (see Chapter 3 for details on DRS, the software used to accomplish this). Based on this replay, the fact that there is an extended pause between the second use of *yeah* and the following utterance from <$1>, supports the claim that this exists as a form of backchanneling behaviour as a pause is used instead of subsequent talk, which would instead make the contribution a turn rather than a backchannel. Conversely, the fact the first *yeah* only displays a small second-long pause before *well I've*

been…, suggests that this is being used as part of a full turn instead. It is important to note, however, that although prosody is emphasised as an invaluable 'clue' for backchannel definition, it is not explored in any great detail in Chapter 5.

Following from these discussions, it is appropriate to question the number of turns contained in the excerpt (Figure 2.4), away from the 'theoretical' total of 10 turns given above. Figure 2.6 redefines the location of turns in the transcript excerpt:

```
    ⌠ <$2> <$=> Oh well I <\$=> I'm just reading things at the moment and
    |        just+
  a ⟨ <$1> Right.
    ⌡ <$2> +kind of vague+
  b   <$1> So what given that the amount of stuff on metaphor is huge?
    ⌠ <$2> Yeah well I've been looking through some of the stuff on scientific
    |        metaphors and+
    | <$1> Uh-huh.
  c ⟨ <$2> +er particularly how they're used for educational purposes in+
    | <$1> Right.
    | <$2> +explaining concepts erm+
    ⌡ <$1> Yeah.
```
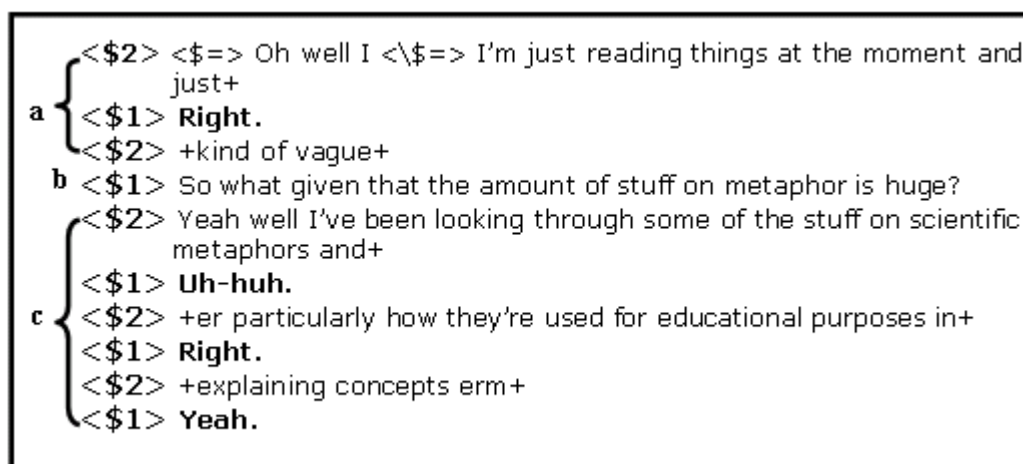
Figure 2.6: Defining turns in a transcript excerpt.

If those items highlighted in Figure 2.5 are taken as backchannels, the first three sequences of talk (*Oh well I* to *kind of vague*) can be classified as one turn, from speaker **<$2>,** marked as **a** on the transcript in Figure 2.6, which is followed by a turn from speaker **<$1>**, the '*metaphor'* question, marked as **b** in Figure 2.6. A final turn is then delivered by speaker **<$2>** in response to the question, taking us to the end of the excerpt (marked as **c** in Figure 2.6). In

short, this excerpt would seem to contain only 3 turns, framed by 4 intermittent listener backchannels.

2.3.1.4. Traditions of backchannel research

The methodological paradigms used when investigating backchannels tend to be motivated by one of two key research traditions in linguistics; Distribution Analysis and Conversation Analysis (Drummond and Hopper, 1993a: 159).

The main concern in Distribution Analysis is to explore the frequency with which specific linguistic phenomena, in this case backchannels, occur across certain contexts. In other words Distribution Analysts seeks to highlight how the usage of backchanneling varies, according to frequency, from one context of conversation or from one speaker to another. This approach was commonly used in early linguistic research (for examples see Duncan, 1972), especially in pre-corpus studies where more primitive techniques for data collection and analysis were available, and involved the manual counting of the frequency items (see section 2.2.2).

In contrast, Conversation Analysts are less concerned with frequency of occurrence of phenomena, instead they are more interested in exploring the specific semantic associations and pragmatic functions that the given behaviours adopt in discourse, and how these behaviours operate in generating sense and meaning. The concern for these researchers, then, is less numerical and more meaning-based.

To a certain extent, this thesis draws on methodological practices used in both of these research paradigms as a means of constructing an integrated and pragmatic CL approach to MM data. Distribution Analysis based

techniques provide an interesting basis for linguistic enquiry by offering a simple, numerical point of comparison between the occurrence of different forms and functional 'types' of spoken and non-verbal backchanneling; through frequency counts and statistical analyses. While the more integrated CA approach adds a more detailed description of the reasons for emergent patterns in the numerical analyses, by providing a pragmatic, contextually-based commentary of what is found.

A combination of these approaches, therefore, allows for the construction of an engaged interpretation of what the differences in the statistical analyses actually mean for the analysis undertaken in this thesis, and for wider explorations of how this behaviour helps to develop meaning in discourse.

2.3.2. Spoken backchannels

2.3.2.1. Forms

On an elementary, lexical level, spoken backchannels can be divided into three major groups; 'simple' (first coined by Oreström, 1983), 'double' and 'complex' (Tottie, 1991: 263, although double and complex forms are jointly referred to as 'series' forms in Oreström, 1983: 121).

Simple forms are brief 'mono or bisyllabic utterances' (Gardner, 2001: 14) comprising single words which generally make up the most frequently used backchannel forms. Examples include *yeah* and *mmm*. Double backchannels comprise a sequence of a specific lexical form which is repeated two or more times, for example, *yeah yeah*. Finally, complex backchannels are composed of 'one of several items from different backchannel categories and/or one of

several open-class lexical items' (Tottie, 1991: 263), such as *yeah….right* or *yeah I know*.

Long, multiple-word, 'complex' backchannels are thought to be particularly common in situations where listeners are not immediately requesting to take the floor, but rather signalling a desire to do so in the near future. Dittman and Llewellyn thus define such phenomena as 'turn requesting backchannels' (1968); 'a way of 'queuing up' or 'negotiating' for the floor, with a function that is 'similar to a raised hand in a classroom' (Oreström, 1983: 124). Given this characteristic, Cutrone notes that, in the case of complex forms particularly, 'sometimes what starts as a backchannel may end up as a turn, if the primary speaker shows no willingness to continue speaking' (2005: 242).

2.3.2.2. Functions

As well as adopting a variety of lexical forms, it is important to emphasise that backchannels are also 'used to achieve a systematically differentiated range of objectives [discourse functions] which, in turn, are specifically consequential for the onward development of the sequences in which they are employed' (Heritage, 1984: 335).

Before this notion of *function* is investigated in more detail, it should be noted that, although it is difficult to 'explain exactly what each [backchanneling] token is used for' every time it is used in conversation (Gardner, 1997b: 12), a wealth of linguistic research exists that conceptualises how specific backchannel forms are *commonly* used, in terms of their discursive functions.

O'Keeffe and Adolphs (2008: 84) provide a good example of a functional coding model that categorises backchannels according to four different sub-groups; Continuers (CON), Convergence tokens (CNV), Engaged Response tokens (ER) and Information Receipt tokens (IR). These broad categories are extended in this thesis (see Chapters 4, 6 and 7) to incorporate non-verbal backchannels in order to create an integrated coding system for labelling all backchannel types, and for influencing explorations of the relationships between the existence and use of spoken *and* non-verbal varieties (similar models are given by Maynard, 1989 and Gardner, 1998):

- CONTINUERS:
  - The most basic form of backchannel, which is used to maintain the flow of discourse, and to provide feedback on how the message is being received.
  - Continuers act as floor-yielding tokens signalling that the addressee is listening, desiring the speaker's floor holding narrative to continue.

- CONVERGENCE TOKENS:
  - Convergence tokens have a 'higher relational value' than continuers, as they are used to mark agreement / convergence.
  - They are used to help maintain good relations, by reinforcing commonality throughout the discourse.

- ENGAGED RESPONSE TOKENS:

  ▪ These are more affective response tokens, communicating emotive signals and opinions to the speaker without taking over the turn.

  ▪ They can highlight, for example, the addressee's anger, shock, surprise, disgust, sympathy, empathy and so on.

- INFORMATION RECEIPT TOKENS:

  ▪ These are highly organised tokens which are associated with asymmetrical discourse, where one speaker has control over the flow of discourse.

  ▪ They are rare in casual conversations in familiar settings.

  ▪ They can assume the role of a discourse marker, signalling the close or shift of a topic (so are usually marked by falling pitch).

These are seen to exist on a functional cline (O'Keeffe and Adolphs, 2008), a 'continuum of facilitative interactional feedback' (Stubbe, 1998b). Positioned at one end are backchannels that are most 'facilitative' (O'Keeffe and Adolphs, 2008: 84), those that are primarily involved with the management of structure and flow of the discourse. These are backchannels with a relatively low lexical or affective content and a 'neutral affect' (Stubbe, 1998a: 258). These backchannels therefore commonly function as CON and CNV tokens.

Schegloff first coined the notion of CON tokens (1982), although Fries similarly talks about 'signals of continued attention' (1952: 49). CON are

typically simple form backchannels, although some double forms of backchannels also function as CON, and are often noted as being the most common function of backchannel behaviour (see Oreström, 1983; Tottie, 1991 and Cutrone, 2005). O'Keeffe and Adolphs (2008) support this claim in a study which explored patterns in the use of backchannels, as evidenced by two 60,000 word spoken language corpora datasets taken from CANCODE and LCIE (the Limerick Corpus of Irish English). This study found that an astounding 97% of the backchannel forms identified functioned as CON tokens, providing minimal feedback to the speaker[10].

Gardner identifies the primary function of a CON as being *prospective* (2001: 16, based on Jefferson, 1984), operating in assisting in the formation of 'bridges between units of talk' (Goodwin, 1986: 209). CON backchannels mark an immediate shift of the floor back to the prior speaker, clearly maintaining the flow of the conversation.

The second function cited here, CNV tokens, is used to describe backchannels that function to signal an acknowledgment of hearing, understanding or agreeing within the conversation, emitting some form of convergence, association and emotional response to the speaker (O'Keeffe and Adolphs, 2008: 77). CNV tokens can be single word items, or can appear as follow-up questions and short statements, and so can feature as double or complex forms of backchannels. CNV backchannels are positioned in close proximity to CON backchannels on the *functional cline*. As although they have a slightly greater lexical content and relational value than CON forms, and are generally more affective, they are still relatively simple forms of backchannel

---

[10] This finding provides the stimulus for premise 4, see Chapter 5 for details.

in comparison to the more complex ER and IR varieties (although this is not true of all forms of CNV tokens).

At the other end of the *functional cline* are backchannels which are used in a more 'relational' way (IR and ER tokens, see O'Keeffe and Adolphs, 2008). They are often used in 'asymmetrical discourse where one speaker has managerial power over the flow of the discourse' (O'Keeffe et al., 2007). These are affective markers in discourse, used to signal a high involvement of the listener, marking 'positive affect or cooperative overlaps' (Stubbe, 1998b). Therefore, they are examples of backchannels with a very 'clear lexical content', away from, the 'empty' CON tokens (Goodwin, 1986: 214).

ER backchannels fulfil one of two key functions in discourse, either in an affective way, to signal 'raised engagement' and express a wide range of opinions and feelings, from anger and shock, to sympathy and empathy, or as a 'turn-claiming signal' (Oreström, 1983: 175). Finally, IR tokens are usually highly organisational and are most commonly evident in professional discourse contexts, thus are quite 'rare in casual conversation' (O'Keeffe and Adolphs, 2008).

2.3.2.3 The relationship between forms and functions
The literature suggests that although there is no specific one-to-one relationship between the lexical form and discursive function of spoken backchannels, in many cases specific forms do have a tendency of adopting one function more frequently than others.

One of the most frequent lexical forms of CON backchannels is the simple response token *mmm*[11] (also spelt as *m* and *mm* in the literature, as seen in Tottie, 1991 and Oreström, 1983). Oreström's study of backchannels in the BNC found that *mmm* occurred in 50% of all cases examined (1983: 131). This result was matched in another, similar study by Oreström (1983, also see Gardner, 1997a, 1997b and O'Keeffe and Adolphs, 2008) in which he surveyed the use of backchannel forms across 10 conversations from the LLC (London-Lund Corpus, another British English corpus) and revealed that *mmm* was used in 50% of all cases, with *yes* being used in 34%, *yeah* in 4%, *mhm* in 4%, and *no* in 3% of all cases.

Gardner describes *mmm* as the 'most minimal response' (1998: 210) in that it contains very little semantic content (see Cathcart et al., 2003: 51). It is generally understood that *mmm* is more likely to be uttered 'jointly with the ongoing speech' than other backchannel forms perhaps are, in other words it is unlikely that the utterance of this response will interrupt a speaker in a conversation (this was the case for half of the *mmm*'s seen in Oreström's study, 1983: 131), making it 'insufficient on its own to do the work of heralding a topic change' (Gardner, 1997a: 135). Consequently, *mmm* is rarely used to function as anything other than a CON in discourse.

As a result of these characteristics, Bublitiz suggests that *mmm* is the most dexterous of backchannel forms, insofar as it may legitimately be placed at any point in discourse (1988)[12]. However, despite this dexterity, it is noted that the response *mmm* is commonly positioned at a TRP as a means of ensuring effective cooperative talk (refer to Grice, 1989), because if they were

---

[11] This finding provides the stimulus for premise 3, see Chapter 5 for details.
[12] This finding provides the stimulus for premise 3, see Chapter 5 for details.

merely placed haphazardly throughout the discourse, it could act as a signal that the hearer is not listening attentively, and it is likely that, after a while, the speaker would detect this (Bublitz, 1988).

The second most common CON backchannel form is *yeah*[13]. Gardner provides extensive explorations of the use of this lexeme (see Gardner 1997a, 1997b, 1998, 2001- he also provides explorations of *mmm, mm hm, mh huh, yeah, oh,* and *alright*). *Yeah* is thought to be more multifunctional than *mmm* (Gardner, 2001: 17, see also Beach, 1993 and Drummond and Hopper, 1993a, 1993b), as while the vast majority of *mmm* forms are examples of backchanneling, with few exceptions, *yeah* is not only used as a backchannel, but commonly comprises part of a full turn, as was evidenced in Figure 2.6.

In a study conducted by Drummond and Hopper (1993a, this study also examined the use of *uh-huh* and *um hm*), it was discovered that in only 22% of all instances, *yeah* was used as a 'minimal' contribution (i.e. backchannel) while it was used as a 'full' turn in 36% of the total number of uses. This stands in stark contrast to *mmm*, as a study by Gardner (a corpus-based study investigating the use of backchannels in 7 hours of Australian English) indicated that of the 700+ instances that *mmm* were used, practically all existed as a free-standing simple backchannel (1997a: 135, also see Beach, 1993 and Gardner, 2001). This suggests that *yeah* has a relatively higher 'degree of *speakership incipiency*' than *mmm* (Drummond and Hopper, 1993a, in support of earlier claims made by Jefferson, 1984).

Yeah also adopts a greater variety of pragmatic discursive functions than *mmm*. Not only can it function as a polar response to a question, it can be

---

[13] This finding provides the stimulus for premise 3, see Chapter 5 for details.

used, as a backchannel, either merely as a minimal CON, or can be engaged to do 'some varying kinds of acknowledging, affirming or agreeing work, as well as showing, for example, surprise, appreciations, assessments and so on' (Gardner, 2001: 34). Consequently, *yeah* is often used as an 'archetypal acknowledgement token in English' (Gardner, 2001: 34), i.e. as a CNV backchannel. Owing to the multi-functional nature of *yeah*, it can sometimes be difficult to determine the function of this response on any given occasion. This matter is discussed in Chapter 4.

Other common backchannels often used as CON tokens, *uh-huh* and *mm hm* (see Schegloff, 1982 and Jefferson, 1984, 1993 for specific studies on these), are more passive than *yeah*, in that they rarely demonstrate speaker incipiency (Drummond and Hopper, 1993a: 158). Drummond and Hopper determined that of all the *mm-hm* and *uh-huh* tokens used in their study, only 5% and 10% were thought to be examples of speaker incipiency, the remainder were backchannels. In contrast, this figure stood at 45% for the *yeah's* examined (Drummond and Hopper, 1993b: 203-4).

The most common forms of CNV tokens are single word backchannels such as *yeah* (see above for details), 'echo questions' like *did you?*, and short statements such as *yeah its pretty sad* (see O'Keeffe and Adolphs, 2008 for more examples). So while CON tokens are most frequently 'simple' form backchannels, CNV tokens are sometimes more structurally complex, mirroring the more affective semantic associations of these terms, in comparison to the CON.

ER tokens, on the other hand, are most commonly series (i.e. complex) backchannels (see Oreström, 1983); consisting of multiple word utterances[14]. These often comprise short statements or repetitions such as *oh really* and *that's nice*, although these tokens can also appear as simple, single word forms such as *excellent* and *absolutely* (O'Keeffe and Adolphs, 2008: 84).

A backchannel frequently functioning as an IR token, which has been extensively explored in linguistic research, is *oh* (see Heritage, 1984, 1998 and Gardner, 2001). Schegloff (1982) identifies that *oh* is often 'followed by further talk' by the listener, and, therefore, does not always act as a backchannel in discourse (Gardner, 2001: 41). Instead *oh* can be used, in cases of speaker incipiency, as the start of a turn, particularly when it is functioning as a discourse marker in the conversation (see Schegloff, 1982 for further details).

Oh, in a similar way to the response token *ah*, commonly functions as either a 'topicalizer' (developing the talk) or as a 'follow-up' (developing or changing the topic of conversation). Therefore it frequently acts as a 'change of state [activity] token' (Heritage, 1984: 307, see also Aijmer, 1987; Stenström, 1987; Heritage, 1998 and Gardner, 2001: 41). This is because it marks the receipt of information, that is, a change of state of knowledge or understanding of the listener, and signals a wish to either change the topic of talk; to mark the end of a story, the end of a conversation (illustrated by Gardner, 1997b: 30) or to project 'a preparedness to shift from recipiency to speakership' (as with *yeah*; see Jefferson, 1984: 200).

---

[14] This finding provides the stimulus for premise 5, see Chapter 5 for details.

The 'topic shift' characteristic possessed by *oh* and *ah* (O'Keeffe and Adolphs, 2008: 86) is shared by other forms of IR backchannels, including the simple forms *right* and *okay,* both of which are common in discourse (see Stenström, 1987 for investigations of *right*, and its marginal derivatives, including *right o, all right, that's right, that's all right* and *it's all right* in a sample of data from the SeU, also see Beach, 1993: 328; Beach, 1995 and Gardner, 1997b: 30 for discussions of *okay*).

Given these comments, we can now appropriately encode the specific functions of the backchannels featured in the transcript except seen in Figures 2.4, 2.5 and 2.6. This is illustrated in Figure 2.7.

```
<$2> <$=> Oh well I <\$=> I'm just reading things at the moment and
     just+
<$1> Right.
<$2> +kind of vague+
<$1> So what given that the amount of stuff on metaphor is huge?
<$2> Yeah well I've been looking through some of the stuff on scientific
     metaphors and+
<$1> Uh-huh.
<$2> +er particularly how they're used for educational purposes in+
<$1> Right.
<$2> +explaining concepts erm+
<$1> Yeah.
```

Figure 2.7: Defining the functions of the backchannels seen in the transcript excerpt taken from the NMMC.

In this figure the two uses of *right* (marked in red) in the transcript excerpt (seen in Figures 2.4, 2.5, and 2.6) appear to be backchannels functioning as IR tokens in that they provide feedback which is highly organisational, marking that information has been received, understood, thus indicating a change/ shift in the listener's understanding. On the other hand, *uh-huh* and

yeah, are functioning as CON backchannels (marked in yellow) as they provide brief, minimal feedback to the speaker which is low in content and affect.

2.3.2.4. Backchanneling in context

As previously identified, 'interpersonal communication does not occur in a vacuum, it takes place in cultural contexts, that is, a system of norms and values' (Myers and Myers, 1973: 215). Therefore, it is necessary to consider the socio-cultural context of talk when outlining how spoken or non-verbal backchannels are used, and to help determine their specific function in the discourse.

A plethora of past studies have sought to explore the difference in the use of specific spoken backchannel forms, their attributed functions and frequencies of use across speakers from different socio-cultural backgrounds (these include studies by White, 1989; Maynard, 1990; Stubbe, 1998a and McCarthy, 2002). In general, these studies indicate that backchanneling is heavily culturally and contextually specific.

Cutrone (2005) carried out such a study, investigating patterns of backchannel use between separate dyads of all-British and all-Japanese speakers. He found that the Japanese speakers generally used slightly more backchannels, and that, in each case, the different groups of speakers used backchannels to fulfil different discursive functions. This result is also supported in similar studies conducted by White, 1989 and Maynard, 1997.

This result is interesting, especially in that many of the studies surveyed thus far (including those by Fries, 1952; Kendon, 1967, 1972; Yngve, 1970;

Duncan, 1972 and Schegloff, 1982), have specifically focused on examining the backchanneling behaviour of American English or Australian English participants. Comparatively, fewer notable studies exist that focus on British English forms, aside from those by Oreström, 1983 and O'Keeffe and Adolphs, 2008. While it is logical to assume there will not be a dramatic difference in patterns of backchannel use between American and British participants, as that between Japanese and English participants, it is misleading to suggest that absolutely no difference exists.

Therefore it is important to acknowledge that, while these studies have revealed some interesting facets of backchanneling use, the results are not necessarily directly transferable to the current study.

To exemplify this point, Tottie conducted a comparative study of backchanneling behaviour between groups of all-American English and all-British English speakers (1991). From an examination of two separate conversations from each group she discovered two key differences between the speakers. Firstly, the average amount of backchannels administered per minute differed dramatically across the groups, with the American speakers backchanneling more frequently (16 backchannels per minute) than the British speakers (5 backchannels per minute). Secondly, she found that the most common lexical forms of these backchannels also differed across the groups, as shown in Figure 2.8.

American English		British English	
Backchannel Form	**Percentage of use**	**Backchannel Form**	**Percentage of use**
yeah	40	yes	44
mhm	34	m	36
hm	11	no	26
right	4	yeah	4
unhhunh / uhuh	4		

Figure 2.8: The percentage of use of the most frequent backchannel forms spoken by groups of American English and British English speakers (results taken from Tottie, 1991).

Figure 2.8 indicates that while *yeah* is only used in 4% of the British English data, it is an overwhelming 40% for the American English data.

Interestingly, the figure of 4% given for the use of *yeah* in British English, as seen here, dramatically contrasts with the results found in a further British English study of backchannels, conducted by O'Keeffe and Adolphs. This study, which compared British English forms to Irish English forms, again identified key differences between the two language varieties. However unlike the example in Figure 2.8, *yeah* was actually established as the most common backchannel form in the British English (as evidenced by an analysis of 60,000 words from the British CANCODE corpus, see O'Keeffe and Adolphs, 2008).

This finding suggests that, as with all corpus-based studies, results yielded from an analysis of a specific corpus dataset are not necessarily consistent across all discourse contexts, or all speakers, despite being representative of the cases that are examined in the given study(ies). Owing to this potential for inconsistent results across socio-cultural contexts, this thesis therefore

focuses only on conversational data elicited from all-British-English speaker dyads (a combination of female-female, male-male and female-male), in an academic discourse context, in order to create, as far as possible, a consistency across the data analysed. Obviously, despite this it is necessary to remember that idiolectic differences, from speaker to speaker can also affect the results generated from analyses, in the same way as such broad socio-cultural categorisations of participants.

Further to the socio-cultural context, it is important to briefly mention that factors such as status, gender (for specific examples of this see: Hirschman, 1974; Thorne and Henley, 1975; Duncan and Fiske, 1977; Maltz and Borker, 1982; Roger and Schumacher, 1983; Blum-Kulka and Olshtain, 1984; Brown and Levinson, 1987; Roger and Nesshover, 1987; Bilous and Krauss, 1988; Dixon and Foster, 1998; Henley and Kramarae, 1991; Kasper, 1995; Mulac and Bradac, 1995; Mulac et al., 1998 and Heinz, 2003) and the relationship between participants involved in a conversation, can also potentially influence the form, frequency and function of backchannel usage in discourse.

These factors are, thus, vital to consider when embarking on investigations of real-life discourse phenomena. Consequently, they are discussed and re-examined in relation to the patterns, results and analyses witnessed in the main study of the thesis (see Chapters 5 and 6 for details).

2.4. Communication 'beyond the text'

2.4.1. Language and gesture

The previous section concentrated on the spoken characteristics of talk, and backchanneling behaviour in particular. This section provides a background to 'non-verbal' features since:

> Languages contain not only words, phrases and sentences but languages also have imagery; they have a global, instantaneous non-compositional component that is as defining as the existence of a language as are the familiar linguistic components. (McNeill et al., 1994: 223)

Discourse, therefore, comprises not only of spoken or vocalised features, but also sequences of non-verbal behaviour (NVB). NVB includes a wide range of phenomena such as hand and arm movements (see Thompson and Massaro, 1986; Rimé and Schiaratura, 1991 and Beattie and Shovelton, 1999 for studies related to these forms of gesture-in-talk), gaze (see Griffin and Bock, 2000 and Cerrato and Skhiri, 2003), body movement, head nods and facial expressions (see Black, 1984; Ekman, 1982; 1997 and Black and Yacoob, 1998).

There is a general consensus that speech and such forms of NVB interact on many levels in discourse. The closeness of the relationship can, however, be widely debated and since 'there is no single theory of non-verbal communication any more than there is a single theory of social behaviour'

(Argyle, 1988: 9), a range of different, sometimes conflicting, opinions and associated theories have emerged over the years.

Many early studies of gesture-in-talk actually worked from the premise that 'the system of gestures is very different in its underlying principles from the system of language' (Chomsky, 1983: 40, also see Dittman, 1960: 341). Maintaining that, since spoken and non-verbal (gestural) aspects of talk differ dramatically in terms of physical manifestation (i.e. form), communication is best conceptualised as being comprised of a variety of different, separate 'channels', with distinct 'spoken' and 'non-verbal' behaviours. As part of this conceptualisation, NVB was thought to 'serve functions totally different from those of language, and perform functions which verbal language is unsuited to perform' (Bateson, 1968: 615).

Therefore, although these theorists (including Dittman, 1960; Chomsky, 1965 and Bateson, 1968) did acknowledge that gestures *are* important to communication, they maintained that they perhaps exist only as a metalinguistic 'paralanguage' (Argyle, 1988: 104).

Other theorists found this perspective to be problematic and, starting with Birdwhistell (1952), believed the relationship between language and gesture to be inherently closer than this (for further studies examining the co-occurrence of speech and body movements from this perspective, see Kendon, 1972, 1980, 1994, 1997; Schegloff 1984; McNeill 1985, 1992; Nobe, 1996 and McClave, 2000). Conversely, it was argued that gestures operate simultaneously (see Brown, 1986: 409) and with 'close synchronicity' to spoken words (Kendon, 1972, also see McNeill, 1985), interacting, and sometimes 'counteracting', with them in discourse (Maynard, 1987: 590).

These theorists postulate that certain forms of NVC (not all, as discussed in section 2.4.2.1) are in fact 'truly part of speech in that they contribute to meaning just as words and phrases do', rather than merely existing as a distinct subsidiary to speech (Bavelas, 1994: 205). This view implies that such gestures can adopt a wide range of different meanings and functions in discourse, and 'can do almost as many things' as spoken language (Streeck, 1993: 297, see also Chalwa and Krauss, 1994: 580). For example, they can be semantically aligned with the abstract or concrete objects and notions expressed in lexis in order to generate meaning, maintain relationships, and to manage and provide structure to discourse.

McNeill extends this idea by proposing that the relationship between language and gesture is in fact so closely entwined that it is erroneous to actually think of communication in terms of containing 'verbal' and 'non-verbal' elements at all, whether these are conceptualised as being distinct or not (1985). He maintains that some gestures are not actually non-*verbal per se*, rather they are non-vocal (see section 2.4.2.2 for more details).

It is argued that this is because both visual and vocal 'signs' witness the same 'computational stage' in talk, the same psychological process prior to production (McNeill, 1992: 30 also in Kendon, 1979, 1990). It is this pre-production configuration that gives gestures the potential to acquire, in a similar way to words, semantic and pragmatic meanings and functions, although they do not necessarily express the exact same thing at a given time in discourse. Therefore, on production, 'information in both communicational channels complement each other in order to convey the full meaning of a

single cognitive representation' (Holler and Beattie, 2003: 81, see also Holler and Beattie, 2004 and Clark and Krych, 2004: 78).

This idea maintains that gestures are distinguishable from the spoken word only in that a difference in the 'psychological actions of speech production' results in a visual rather than a vocalised sign being produced (McNeill, 1992: 30). Thus, by reinforcing the 'separate channels' approach we are inherently modelling the 'wrong thing' (Bavelas, 1994: 205), neglecting to prioritise the importance of the 'linguistic function' in the conceptualisation of gesture-in-talk, giving precedence to the 'physical source' instead (Bavelas, 1994: 205). This shows a preoccupation with the fact there are innate discrepancies between vocalised and optical signals in talk, owing to differences in the physical manifestation of these signs. Non-verbal behaviour is effectively 'continuous', whereas speech is discontinuous (see Dittman, 1960), as 'even if we are asleep our bodies still emit non-verbal messages' (Richmond et al., 1991: 5), simply because 'there is no such thing as non-behaviour or, to put it more simply: one cannot *not* behave' (Watzlawick et al., 1967: 48).

In contrast to early ideals, although these later theorists acknowledge that speech and gesture *are* physically and semiotically different from each other (i.e. manifested using different 'signs', see McNeill, 1985), they believe these behaviours to be more than 'just movements' (McNeill, 1992: 105), but having a function which is inherently 'linguistic' (Bavelas, 1994: 202).

This thesis, as a linguistic study of gesture-in-talk, is effectively interested in trying to reveal the specific *meaning function* of sequences of movements (i.e. backchanneling head nods). Therefore, the study is concerned with 'how

gestures communicate' (Bavelas, 1994: 201), and the ways in which they interact with specific features of talk in order to achieve this. Therefore, the views of the later theorists are supported here, working on the assumption that a strong relationship does exist *between* spoken and non-verbal forms of backchanneling phenomena, rather than the separate channels theory. Nevertheless, the terms 'spoken' (rather than verbal) and 'non-verbal' are used throughout, simply for ease of reference and to create a consistency in terminology, although the shortcomings of perhaps using such terms (McNeill, 1985) are acknowledged.

2.4.2. Defining non-verbal behaviour

2.4.2.1. NVB Vs NVC

In line with this argument, it is necessary to state that although by definition NVB includes all forms of non-spoken human behaviours that 'have the potential for forming communicative messages' (Richmond at el., 1991: 7), as a linguistic study of such behaviours the current thesis is only concerned with a particular sub-set of these. These are behaviours that adequately *fulfil* this potential, those that are deemed to have some sort of significance or meaning, in talk. So, the focus is specifically on how individuals both 'give' and 'give off' information in interaction using these movements in order to generate meaning (Goffman 1963, cited in Kendon, 1997: 117).

Historically, gestures with a 'potential' to communicate were included under the terms 'kinesic behaviour' (the study of which is known as kinemics, first coined by Birdwhistell, 1952) and 'expressive movement' (see Davis,

1979: 54). A more common and current term is 'non-verbal communication' (NVC).

It should be noted that it is difficult to explicitly describe the differences between NVB and NVC, as both are essentially 'heuristic units' of abstract behaviour, rather than bound concrete or 'static units' (paraphrased from Norris, 2004: 12). Therefore, paradoxically, these phenomena do not lend themselves to rigid definitions. However, there are some fundamental, theoretical differences between these behaviours, and before specific forms of NVC are investigated further, it is useful to underline and model these, as shown in. Figure 2.9 (Richmond et al., 1991: 8):

RECEIVER

	1	2
3	**NVC**	**NVB**
4	**NVC**	**NVB**

SOURCE

1 = Interprets behaviour as message
2 = Does not interpret behaviour as message
3 = Behaves to send message
4 = Behaves with no intent to send Message
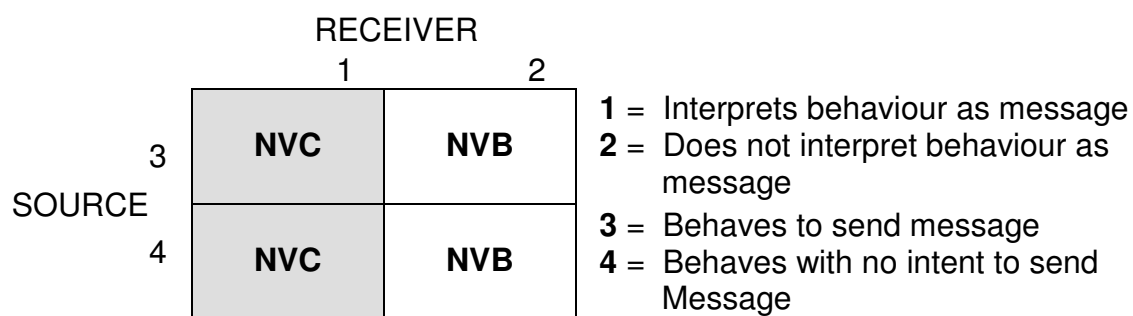
Figure 2.9: The differences between Non-Verbal Behaviour (NVB) and Non-Verbal Communication (NVC).

This figure illustrates the key difference between NVB and NVC. NVC relies on the presence of another party in talk, whereas NVB does not given that NVB is continuous regardless of who is or is not present. In other words, NVC, first and foremost, comprises individual or sequences of discrete and structured gestural episodes which communicate messages *between* 2 or

more individuals involved in a conversation (conceptualised as the 'source' and 'receiver' in figure 2.9).

If the receiver interprets given 'behaviour[s] as a message and attributes meaning to that message' (Richmond et al., 1991: 7), the behaviour is best defined as a form of NVC rather than NVB (see column 1). This is regardless of whether the source intends to send a message (i.e. consciously gestures to the receiver, column 1, 3), or whether there is no such specific intention (i.e. with no conscious intent, column 1, 4). Conversely, instances where a gesture is made, but is neither intended (by the source), nor interpreted (by the receiver) as a message are seen as examples of NVB rather than NVC (see column 2, 4). This is also true of instances where the source intends to send a message to the receiver without the receiver acknowledging that a message has been sent (i.e. the message is not received). Therefore the listener, the 'receiver', has a crucial role in defining NVC in talk.

The relative success of whether a behaviour is, firstly, interpreted as a message, that is, whether it is NVC or NVB, and secondly, whether it generates the same meaning as the source perhaps intended, is highly variable in discourse, although, in the case of 1,4 in Figure 2.9, this intent is not always present as the use of many forms of NVC is impulsive rather than consciously delivered. This is because the meaning attributed to a given gesture or sequence of gestures is, as discussed below, not necessarily discontinuous.

2.4.2.2. The continuum of gesture-in-talk

While in spoken language individual words (parts) are combined to create sentences (the whole), with the individual parts determining the meaning of the whole, that is, lexis are sequentially structured in talk, with NVC it is the complete gesture (*global*) that determines the meaning of the individual parts (McNeill, 1992: 19). To this effect, gestures are 'non-combinatoric' (McNeill, 1992: 20-21), as the meaning of a gesture is globally defined, 'a complete expression of meaning unto itself' (McNeill, 1992: 21), rather than being a sum of each of its individual parts.

Although there are exceptions to this, such gestures are more flexible than lexis, as the combination of numerous sequences of individual gestures can create various different structures of meaning. McNeill, (1992: 184) defines this characteristic as *synthetic*. By contrast, lexis are organised in more specific and structured ways insofar as the process of, for example, adding words to other words (i.e. a prefix to a word) or adding a subordinate clause to a main clause, is bound to a certain extent by predefined rules of grammar (see Norris, 2004: 2). Thus, in short, lexis lacks the global-synthetic characteristic that is inherent in NVC, this is the characteristic which creates discontinuity in the meaning attributed to such forms of behaviour.

The global-synthetic nature of NVC means that there are potentially an infinite number of different gesture sequences that can operate in conversation. This makes the classification, interpretation and exploration of these movements challenging. However, a number of comprehensive models exist which aim to assist in this classification. The most widely used of these is presented as a continuum of NVC, developed by McNeill (1985). This is

depicted in Figure 2.10 (McNeill, 1992: 37, illustration taken from Kendon, 1997).

| Gesticulation | → | Language-like gestures | → | Pantomimes | → | Emblems | → | Sign Language |

Figure 2.10: Kendon's continuum of NVC (based on McNeill, 1992).

Positioned at the right side of the continuum is sign language. Signed languages predominantly consist of pre-defined, highly structured linguistic codes which concentrate on the use of '(primarily) optical signals' (McNeill, 1985: 351), rather than '(primarily) acoustic signals' as a means of communication, although acoustic signals *are* integrated into some signed languages (see Vermeerbergen, 2006 for a review of key studies of signed languages). In other words, sign-languages use their own self contained and conventionalised symbolic lexicon as a means to communicate (Kendon, 1997) and, thus, do not theoretically rely on the co-occurrence of speech.

Similarly, emblems and to a certain extent pantomimes, are also highly conventionalised gestures (Kendon, 1992) which can be used and successfully interpreted in the absence of speech. Goldin-Meadow suggests that emblems are most the salient forms of gestures for speakers and listeners in non-signed environments (1999: 419), because they consist of fixed, 'standard sequences of human behaviour' (Argyle 1988: 142), which are often considered to be *ritualistic*. Examples of emblems include the 'thumbs-up' sign and the 'ok' sign. The meaning of an emblem is commonly specific to the socio-cultural context in which it is used (Holler and Beattie, 2002). So

although emblems are commonplace in many different cultures, the meaning attributed to these signs is not necessarily consistent across them (McNeill, 1985: 351).

This characteristic is explored extensively in the work of the etiologist Desmond Morris. Morris compares the meaning of the emblematic 'thumbs-up' gesture in various different socio-cultural contexts, deriving a range of meanings, as seen in Figure 2.11 (Morris et al., 1979). This idea is also explored in Streeck (1993) who undertakes a cross-cultural and cross linguistic study of the organisation of speech and gesture in face-to-face interaction.

'Meaning'	Frequency of meaning (from a total of 1200 instances)
O.K.	738
One	40
Sexual Insult	36
Hitch-hike	30
Directional	14
Others	24
Not Used	318

Figure 2.11: A table to show a variety of semantic associations of the 'thumbs-up' gesture, based on 1200 participants across 40 different locations around the world (taken from Morris et al., 1979).

The figure highlights that, although the thumbs-up symbol means 'ok' in the majority of the instances presented, in others it can be accorded an altogether different, and sometimes incongruous, meaning. Therefore,

meaning is very much dependant on *where*, *when* and *whom* is using and/or interpreting the gesture. Consequently, the use of the thumbs-up symbol bears no relationship to the pragmatic and semantic content of speech with which it may or may not co-occur. Instead, the specific meaning of this conventionalised gesture is often 'learned as [a] separate symbol' by a person during the 'process of socialisation' (McNeill, 1985: 351).

The next form of NVC seen in Figure 2.10 is the pantomime. Pantomimes are the 'larger than life' facial expressions or bodily movements, such as accentuated smiles or frowns, which can be used in conjunction with speech to express a range of moods and emotions. Again, as with sign-language and emblems, these highly structured gestures are socio-culturally and/or context specific. Pantomimes have attributed meanings which are not strictly reliant on the spoken aspects of concurrent talk; insofar as speech is not necessarily 'obligatory' (Kendon, 1997) for generating the meaning of these gestures, because, effectively, they can be used in silence as a substitute for speech.

On the other hand, gesticulation and language-like gestures have a closer relationship to speech that occurs or co-occurs with them. They do not adopt a fixed or standardised movement structure in conversation (for example, it is unlikely that two hand motions will be exactly the same), as they are spontaneous and transient (Bavelas, 1994: 209) forms of NVC. Therefore, these forms do not have a pre-determined, easy to define one-to-one relationship with meaning, as shown in McNeill's notion of the *global-synthetic* in section 2.4.1, instead their attributed meaning is generated in the context of the lexical environment in which they are used. Consequently, gesticulation and language-like gestures are difficult to interpret without speech.

The key difference between these two forms of gesture is, however, that language-like gestures are most frequently integrated with grammatical features of speech, a specific syntactic structure, for example, they may be used to substitute for a particular adjective in an explanation. Gesticulations, on the other hand, are described as being more free-form insofar as they are not always used as direct substitutes for specific grammatical or syntactic features of talk.

Forms of gesticulation allow speakers, in this case, although this is also true for listeners, to express extra information about abstract concepts, opinions and emotions that are not necessarily 'readily be expressed in words' (Argyle, 1998: 141, also see Wilcox, 2004: 525; McNeill, 1985: 360).

The use of gesticulation also assists in maintaining the 'relationship or common' part of a message (Noller, 1984: 7), functioning both socially to sustain relationships between participants within discourse, and pragmatically as a tool of managing discourse and discourse structure, as well as carrying semantic content.

Richmond et al. outline six key discursive functions of gesticulation, as follows (1991: 8):

- Complementing
- Contradicting
- Repeating
- Regulating (regulating the flow, e.g. looking away)
- Substituting
- Accenting (emphasizing a spoken message)

It is possible for a given sequence of gesticulation to adopt more than one function in discourse at a given time as these classifications are not strictly taxonomic (Bavelas, 1994: 204). For example, the function of a given 'repeated or extended gesture' may not only be used to 'depict information but also [to] convey emphasis or seek a response' (Bavelas, 1994: 204). Consequently, such a gesture would be seen as fulfilling both the 'accenting' and 'repeating' (as a request for clarification) functions at that given point in the conversation.

A variety of different matrices exist which that attempt to sub-characterise forms of gesticulation, since this is the most prolifically researched form of NVC. McNeill provides the most comprehensive and widely used classification system for this, as follows (1992, 1985, alterative categorisation schemes can be found in Ekman and Friesen, 1969; McNeill et al., 1994: 224; Richmond et al., 1991: 57 and Kendon, 1997):

        A. Iconics

        B. Metaphorics

        C. Beats

        D. Cohesives

        E. Deictics

The above five categories of gesticulation provide a useful classification system based on *how* these various forms of gesticulation communicate; how they function to add meaning to a spoken message, this is true of the majority of research into NVC, (see Bavelas, 1994: 201 for further details).

Beyond this, it is difficult to provide clear definitive blueprints on how to reveal exactly *what* specific meaning and/or discursive function is associated with a given gesture or sequence of gestures, as, again, the meaning generated by each of these forms of gesticulation is highly dependent not only on the shape or physical manifestation of the gesture, but on language that may or may not accompany it, together with the patterns of language and gesture used by the other participant(s) in the conversation. This is also true, to some extent, for other forms of gesture-in-talk featured in Kendon's continuum.

Given this, gesticulation is often described as being the *most* idiosyncratic type of gesture-in-talk. So, in short, 'it makes no more sense to suggest that the [specific] linguistic function' or the associated meaning of such forms of gesticulation 'is determined by its physical manifestation than to suggest that the function of a word is determined by the letter it begins with or the phoneme it contains' (Bavelas, 1994: 205). As with the spoken word, while some forms of gesture-in-talk 'have invariant meanings, others [only] have a probability of meaning something' (Argyle, 1988: 6).

The different forms of gesticulation presented by McNeill are hierarchically structured in a similar way to the continuum of gesticulation shown in Figure 2.10 (from A to E). The types of gesticulation with the closest relationship to the actual semantic content of concurrent speech are featured at the top of this list (i.e. iconics and metaphorics). Those which are least semantically tied are located at the bottom (i.e. deictics). These five forms of gesticulation are also listed in the order of how consciously they are utilised in discourse, with iconics being used with the least conscious intent and deictics the most.

Conscious gestures, and those used in a more semi-conscious or unconscious way, are explored more fully by Patton and Giffin (1981: 217).

The category of 'iconic' gestures was first introduced by McNeill in 1985 and is the most widely researched form of gesticulation (Beattie and Shovelton, 2002 and Holler and Beattie, 2002). These are seen to be the form which is the most culturally and contextually tied of all types of gesticulation, an early study supporting this claim was undertaken by Efron, in 1941 (also see Beattie and Aboudan, 1994).

There is a wide range of different, 'complex and often elaborate' (Beattie and Shovelton, 1999: 455) forms of iconic gestures in discourse, many of which are 'associated with different properties of the talk' (Beattie and Shovelton, 2002: 415, McNeill describes iconics as being 'multifunctional' gestures, 1985, also see Beattie and Shovelton, 1999, 2002 for further information). However, they are generally all used to display 'concrete aspects of the scene or event [action] being concurrently described in speech' (McNeill et al., 1992: 224, also see Hadar who refers to iconic gestures as 'lexical movements', 1997: 89). Furthermore, the actual movements enacted by this type of gesticulation also help to illustrate '*how* the action is being accomplished' (Holler and Beattie, 2002: 33).

The most frequently explored iconics are sequences of spontaneous hand movements made by speakers. These studies commonly examine the relationship between these hand movements and some other aspect of communication, such as the direction of gaze for example (as featured in studies by Ekman and Friesen, 1969; Kendon, 1972, 1980, 1982, 1983;

Argyle, 1988; McNeill, 1985, 1992; Streeck, 1993, 1994; Chalwa and Krauss, 1994; Beattie and Shovelton, 2002 and Griffin, 2004).

Most iconic hand gestures are considered to have three main phases of movement (termed the *gesture phase* by Kendon, 1987: 77). The first is the '*preparatory*' phase which exists before the motion occurs, in preparation of the subsequent movement (McNeill, 1992: 12). The crux/ nucleus of the gesture (Kendon, 1987: 77) is described as the '*stroke*' phase (McNeill, 1992: 12), which comprises 'some definite form and enhanced dynamic qualities' (Kendon, 1987: 77). This is the most visible or emphatic part of the gesture, McNeill (1979) refers to this as the *'peak'* of the gesture, while Schegloff calls it the gesture's *'accent'* (1984: 280). The stroke is finally followed by the '*retraction*' phase (McNeill, 1992: 12) which functions to 'either move the limb back to its rest position or reposition it for the beginning of a new gesture phase' (Kendon, 1987: 77). The three phases of iconic gestures involve movements that are not restricted to a specific direction or sequence of rotations, nor are they restricted to a single *stroke*, *reparatory* or *retraction* phase, as a sequence of these may be considered to be part of the same *global* gesture (McNeill, 1992: 19).

It is the *stroke* phase that is most closely integrated with the concurrent speech. The literature suggests that 'the preparation of the gesture precedes the exact lexical units to which it is tied and the stroke often falls on the last accented syllable *prior* to the speakers affiliate' (Schegloff, 1984: 280, also highlighted by McNeill, 1992: 25-26, Streeck, 1993: 280). To this extent, the 'phrasal structure of speech' can be seen to be closely aligned to the 'phrasal structure of gesticulation' for iconics (Kendon, 1987: 77, also shown in

Streeck, 1994: 280 and Kendon et al., 1976). This is the case both *temporally* and *structurally*. Given this, iconics have an important role in conversational management (Kendon, 1997: 111), working with the lexical content to, for example, signal topic and turn shifts following the retraction phase (refer to discussions in respect of TRPs in Section 2.3.1.3).

However, since iconics are spontaneous, the extent to which they comply with this structural organisation and the notion of 'semantic synchronicity' is highly variable (McNeill, 1992: 27). For example, in situations where it is difficult to separate individual or sequences of gestures, defining where they start and stop, it is problematic to determine to what extent this semantic synchronicity is achieved. Therefore, iconic gesture sequences often 'correspond to more than one clause' or correspond simply to 'pauses' and breaks in talk (McNeill, 1992: 27). On these occasions, the gestures still correspond to the phrasal structure of conversation, but they do not necessarily overlap in a strict one-to-one fashion.

Metaphoric gestures are similar to iconics in that they are also closely linked to the semantic content of the speech, adopting the same basic '*gesture phase*' in talk, that is, consisting of the preparatory, stroke and retraction phase. Metaphoric gestures can include gesticulatory movements of the hands, but also the head, arms, torso and certain sequences of proxemic movement.

These gestures tend to be used in 'parallel to sentences with abstract meanings' (McNeill, 1985: 356), instead of being performed with reference to the concrete content expressed in talk, such as being used to signal particular event or objects (as with iconics). In other words metaphorics more frequently

refer to emotions or abstract concepts. Therefore, metaphorics are often used to embody a deeper meaning than the lexis reveals, and, thus, are seen to both relate to, and represent, the context and content of the concurrent speech on a 'meta-level' (McNeill et al., 1994: 224). Metaphoric gestures are least frequently explored in the gesture research literature because their abstract nature makes their meanings somewhat intangible, so difficult to derive and freely interrogate.

The focus of much of the attention into beat gestures has involved the exploration of hand movements, although beat gestures are not always only restricted to these. Beats are 'baton'-like gestures (a term coined by Efron, 1941, also used by Ekman and Friesen, 1969 and Richmond et al., 1991) which are simple, repetitive movements generally comprising a two movement phase. This consists of either an up-down sequence or an in-out sequence (McNeill, 1992: 15, commenting on beat-like hand movements in particular). As a result of this basic kinesic structure, beats are also known as *motor movements* in the literature (Hadar, 1997: 89).

Beat gestures are also described as being 'abstract indicators' (McNeill, 1985: 356), which are coordinated with speech prosody and intonation in talk (as explored by Bolinger 1986: 195; McNeill, et al., 1994: 224; McNeill, 1992; Haiman, 1998; Holler and Beattie, 2002 and Wilcox, 2004). However, unlike the other forms of gesticulation discussed thus far, they are seen to be only tenuously related to the semantic properties of talk. While they can add, for example, emphasis to lexis, they are not semantically *marked per se*, as they have only limited propositional content, unlike iconic and metaphoric gestures. So instead of adding to, or reinforcing propositional content, beats often

function to maintain relationships (McNeill, 1985: 359) and/or are used for general conversational management. Their rhythmic nature is likened to a musical score; maintaining 'flow' in discourse until the speaker desires to give up the turn, at which point the gestural 'beat' is terminated (Chalwa and Krauss, 1994: 583).

The notion of 'cohesive' gestures was developed by McNeill in 1992, building on Halliday and Hasan's discussions of cohesion in speech (1976). Cohesive gestures usually consist of expressive hand movements that 'tie together thematically related by temporally separated parts of discourse' (McNeill, 1992: 16). This is achieved through the use of a set of repeated or similar gestures throughout a conversation as a means of signalling this idea of 'continuation' (McNeill, 1992: 17). Although these gestures do not specifically *relate* to the semantic content of lexis, they instead *work with* semantic content to create cohesion in the discourse. This attribute justifies why they are considered as a form of gesticulation.

Finally, deictics are known as a form of 'kinetograph' in speech (Richmond et al., 1991: 58). These are the most conscious form of gesticulation, which are the least related to the semantic content of concurrent speech. Deictics are, instead, closely related to the conversational *context* of the talk, the physical, conversational (including the relationship between participants involved) and gestural 'space' in which they are enacted (McNeill et al., 1994: 225). Therefore, they are used, mainly though the act of pointing, as a form of spatial reference to 'illustrate location' (Richmond et al., 1991: 58); referring to actual, physical, objects located within the conversational space, or to more

abstract 'imagined objects or locations' and concepts that are constructed and discussed in talk (Holler and Beattie, 2002: 31).

2.4.2.3. Positioning head nods on the continuum of gesture-in-talk

The question to ask, then, is where backchanneling head nods are best positioned within these McNeill models? Traditionally, the physical manifestation of a head nod, whether backchanneling or not, *is* seen as a highly conventionalised movement. Nods are conceptualised as involving a two-step motion, an 'up and down movement, when the head is rhythmically raised and lowered' (Kapoor and Picard, 2001: 3). The most standardised forms of head-nod behaviour are commonly seen to have a purely semantic function; acting as a direct response to a polar question issued by an interlocutor (i.e. nod = yes). For such nods, there is a one-to-one relationship between the nod and its associated meaning. While these nods are semantically tied (similar to gestures on the left side of the continuum), insofar as their relevance is reliant on the speech that precedes it, the interlocutor's question for example, the highly conventionalised use of this gesture may be likened with emblematic forms of gesture-in-talk, rather than with the more spontaneous forms at the other end of the continuum (Efron, 1941 and Ekman and Friesen, 1969 refer to head nods and shakes as emblems).

Yet this is not *always* true of all head nods. Instead, they can adopt a more 'complex movement in which two or all three movement patterns overlap' (Norris, 2004: 33). Consequently, it is often 'difficult to isolate the single movements' of head nods and head nod sequences (Cerrato and Skhiri, 2003: 252), and to accurately associate these particular forms with a specific

role or meaning. Nods of the head can effectively fulfil a range of *different* 'semantic, narrative, cognitive and interactive' (McClave, 2000: 876) functions in discourse, i.e. functions 'beyond affirmation and negation' (McClave 2000: 862).

Maynard conducted a study which sought to outline some of the key functions adopted by head nods, through an observation of the behaviours of six dyadic conversations involving Japanese speakers of English (1987, 1997). He outlined seven key functions as a result of this study, relevant to nods generated by both the speaker and listener in discourse.

Although these are based on Japanese speakers, results from studies of British and American speakers of English suggest that similar functions are adopted by head nods across a range of cultures (see, amongst others, studies by Tao and Thompson, 1991; Feke, 2003 and Heinz, 2003). The key difference, however, is in the *ways* in which these functions are used in different cultures, rather than the meanings derived from them, unlike common forms of emblematic gestures. In other words it is important to regard their frequency and location in discourse, and the specific patterns of physical manifestation of the head nods associated with them (see Cerrato and Skhiri, 2003 and Norris, 2004). Given this, it can be premised that this categorisation scheme will act as a useful benchmark for explorations conducted in this study of British English discourse, in spite of the fact it was not designed to model this language variety specifically. The functions are as follows (Maynard, 1987: 589, also see Maynard, 1997; McClave, 2000; Kapoor and Picard, 2001:1 and Cerrato and Skhiri, 2003 for similar models/ categorisation schemes):

- Backchannel continuer on part of the listener.

- Turn-taking period filler on part of the listener.

- Clause boundary and turn end marker on the part of the speaker.

- Turn transition period filler on the part of the speaker.

- Emphasis on the part of the speaker: (Norris, 2004: 34 suggests that 'we can often determine the strength of the message by the number of times that a person shakes or nods the head').

- Affirmation on the part of the speaker.

- (Pre) turn claim on part of the speaker.

This list again indicates that head nods can carry propositional content beyond the polar response of *yes* (so a head nod ≠ yes). This model, as with the O'Keeffe and Adolphs model (2008) for spoken backchannels, suggests that nods can function in a range of ways, from continuing the flow of discourse or marking agreement (convergence) to acting as more engaged indicators of feedback, and for early signs of movements for the floor.

In effect, regardless of the nature of the sign, spoken and non-verbal backchannels, unsurprisingly, are purported to have the potential for adopting the same basic functions in discourse. The problem is, however, that there is no indication of how given functions relate to the physical manifestation, the movement structure, of nods. While the O'Keeffe and Adolphs model indicated that, for example, spoken forms such as *mmm* has a tendency to function as a CON token, and *right* as a IR token, there is no approximate taxonomy for head nod classification in existence.

The literature also suggests that backchanneling nods, in a similar way to spoken forms of this phenomenon, 'may simultaneously coordinate speech production, mark structural boundaries and regulate turn-taking' (McClave, 2000: 857, also see Birdwhistell, 1970: 103). So, backchanneling nods can help to manage talk and have a close relationship to turns. Head nods can also be used in an interactive way to maintain relationships in conversation and to help structure, manage and 'regulate' interaction (McClave, 2000: 855, a similar range of head nod functions is shown in Maynard, 1987: 589), and provide feedback, as is the case of nods functioning as backchannel continuers, which function to show 'continuation of contact, perception and understanding' (Cerrato and Skhiri, 2003: 256).

Moreover, it is thought that the discursive function of a backchanneling nod is also closely tied to the semantic content of the concurrent discourse. The link between head nods and lexical units was first proposed by Kendon, (1972: 195), also see Goldin-Meadow, (1999: 425). They can be synchronised with linguistic units in speech, and are, therefore, dependant on lexical counterparts to derive the specific meaning function of the gesture.

It is thought that the *stroke* (this can be difficult to freely define, as explored in Chapters 3 and 4) of backchanneling nods are also often aligned with phonetic patterning, specific word forms, ideas or concepts, or specific semantic ideas or relevant positions in the talk (see Kendon, 1972; McClave, 2000; Cerrato and Skhiri, 2003 and Blache et al., 2008) of either the speaker *or* the listener, as is the case of those which co-occur with spoken backchannel forms. More specifically, in a study of backchanneling head nods, Blache et al. summarise that non-verbal similar to spoken,

backchannels, also commonly appear 'after nouns, verbs and adverbs, but not after connectors or linking words between two conversational units' in talk (Blache et al., 2008: 114). This study also discovered that basic patterning and positioning of non-verbal backchannels is generally seen 'to be delayed as compared to the vocal ones' (Blache et al., 2008: 114, also see Dittman and Llewellyn, 1968).

Furthermore, Blache et al's study suggested that a key difference between spoken and non-verbal backchannels is that the nods are less frequently used 'in places of possible turn exchange' than spoken forms, so are used at some 'completion points', although are, on the whole, less often used at TRPs (2008: 114). However, unlike spoken forms, backchanneling nods often occur after specific 'accentual phrases' and 'intonational phrases', whereas spoken ones only occur after 'intonational phrases' (Blache et al., 2008: 114).

Given the various behavioural characteristics, that is, the ability to fulfil a range of roles in discourse, and to adopt a variable movement structure, it is difficult to position backchanneling nods in McNeill's conceptual models. In terms of the continuum, while the two-phase movement structure is typically associated with emblematic forms of gesture, the pragmatic and semantic complexity of these phenomena means that they also have strong associations with gestures at the *gesticulation* end of the continuum. Therefore, it possible to sub-categorise these behaviours either as forms of iconics, metaphoric *or* perhaps beat gestures, depending on which function(s) they are used to fulfil at a given point in discourse.

2.4.2.4. An overview of head nod research

McClave contends that two perspectives for head nod research traditionally exist. These are as follows (2000: 856):

- Their role in speech production.
- Their communicative functions.

Studies of the first type are undertaken from a predominantly physiological (kinemic) perspective, exploring *how*, *where* and *when* nods are performed as part of the physical process of nod production (for examples see Dobrogaev, 1929; Frey et al., 1983; Rimé, 1982; Hadar et al., 1985 and Hadar, 1997). Studies of this nature are common beyond the field of linguistic research.

For example, there is a wealth of research which explore the physical generation and pragmatic qualities of head nods for HCI purposes (see Kapoor and Picard, 2001), thus examining head nods from a computer-science vantage point (also see Hadar et al., 1985 and Sidner et al, 2006), with the view to constructing, for example, real-time models of these behaviours, avatars and service robots. Although important, as will be explored further in Chapter 3, these studies commonly explore interactions between a person and computer (HCI), utilising scripted speech in laboratory conditions to re-create innate head nod use, rather than using naturally occurring examples (see studies cited in Chapters 3, 4 and 6, in addition to studies by Altorfer et al., 2000; Davis and Vaks, 2001; El Kaliouby and Robinson, 2004 and Grönqvist, 2004).

Studies of the second type are more linguistic-functional, and seek to investigate the function of nods in the communicative process. In other words, they examine the interactive properties of nodding; the *what for* of head nods. An example of this, conducted by Cerrato and Skhiri (2003), aimed to observe head nods movements and gaze related to turn-taking and feedback in discourse. During this study, it was discovered that the most common communicative functions adopted by this behaviour were those 'showing continuation of contact, perception and understanding of the message' (Cerrato and Skhiri, 2003: 256). McClave also conducted a study which explored the functions of head nods, observing behaviours exhibited by conversations between dyads of male-male and female-female Americans between the ages of 24-37. From an analysis of this data, McClave provides a comprehensive list of all the functions and types of head nods seen, ranging from the semantic, narrative, cognitive and interactive (2000: 876, in a similar vein to Maynard's list presented in section 2.4.2.3).

Although there are many studies in both of the above fields, there is a comparative paucity of research that adopts a more linguistic functional-kinesic approach, aligning the *how*, *where* and *when* of head nod activity with the *what for*; the semantic, conversational and pragmatic functions of these behaviours, and how these different characteristics combine to generate meaning in discourse (i.e. the main concern of this thesis). Although spoken forms of backchannels, such as *mmm* and *yeah,* and their associated functions have been widely researched by the linguistic community (see, for example, O'Keeffe and Adolphs, 2008), there is limited detailed linguistic research into the various forms and discourse functions of head nod

behaviour, backchanneling nods in particular. Similarly, there is a lack of research that details the closeness of the relationship between such phenomena, first highlighted by Kendon in 1972, and verbalised elements of backchanneling phenomena.

This void in the research can be attributed, primarily, to a range of practical problems associated with physically capturing and representing gesture. Although spoken discourse is relatively simple to record, quantify and observe and such has been undertaken for decades, this is not the case for head nods. The interrogation of non-verbal behaviours involves more technologically sophisticated techniques for the processes of, for example, data collection and quantification. To gain a better understanding of the nature of gesticulation, as seen with spoken backchannels, it is vital to be able to explore not only the physical manifestation of the gesture, but also everything else that is occurring around them, that is, contextual and co-textual features. In order to achieve this, it is necessary to have adequate systems for 'reading' these gestures in context of, and in relation to, the language spoken by each participant involved in the conversation(s) (Goldin-Meadow, 1999: 425). Given that 3rd generation corpora (Section 2.2.2) present episodes of real-life discourse as text, there is a limit to which gesture-in-talk can be *read* using current CL techniques.

The integration of multiple modes of information, as seen in MM linguistic corpora, provides a more complex landscape for exploring elements of data, far beyond that offered by widely used mono-modal corpora. Thus, this enables the examination of patterns of backchannel behaviour use across the 'modes', from multiple perspectives; from form and frequency through to role

and linguistic function. Therefore, in theory, MM corpora will help bridge the gaps in the knowledge of these behaviours, gaps which have been shown to exist in this chapter. Working on this premise, traditional corpus-based methodologies and approaches are adapted in this thesis as a means for providing the facilities for analysing, both qualitatively and quantitatively, patterns in the (co)occurrence of spoken and non-verbal backchannels in videoed dyadic conversations. This will allow the following question to be investigated fully:

What are the roles, forms and functions of non-verbal and spoken backchanneling behaviour in real-life, naturally occurring discourse, and what is the relationship between them?

2.5. Summary

This chapter has provided an overview of key research into NVC, drawing on theories and models of communication from a range of academic disciplines. More specifically, it has also focused on providing an extensive summary of research into communicative feedback, that is, backchanneling phenomena.

The chapter has outlined research into spoken backchannels, describing the various forms, roles and functions of this phenomenon. It has also provided an overview of the comparatively limited research into the physical and functional roles and relevance of non-verbal backchanneling behaviours.

The necessity for more detailed investigations which specifically explore backchanneling nods and their relationship to spoken forms was then

proposed; studies which examine the functional relevance of this behaviour in discourse. This matter will be discussed further in Chapters 4, 6 and 7.

The chapter has also provided a critical review of current CL-based approaches, examining their abilities in allowing investigations of discourse beyond the text (i.e. to explore the use of backchanneling head nods). It has emphasised the necessity for a new refined approach to be developed to allow the user to investigate specific linguistic enquiries in emergent MM corpora.

This notion of the MM CL approach will be further discussed in Chapter 3 exploring, in further detail, the technical, ethical, practical problems and considerations faced in the development and exploration of 4th generation corpora. It also lays the foundations for the MM CL approach that will be utilised as part of the analysis undertaken in Chapters 5 and 6.

Chapter 3: Multi-Modal Corpus Design Methodology

3.1. Introduction

This chapter provides an overview of the principal methodological challenges and considerations faced in the design and construction of 4th generation MM corpora for linguistic analysis. The chapter focuses on investigating the following:

1. How MM linguistic data is collected, encoded, arranged and presented for use.

2. The theoretical, ethical, practical and technological constraints faced during each of the processes identified in (1).

3. How a MM corpus is accessed and used by the linguist, and how subsequent analyses of the data can be undertaken. In other words, it examines how the records of conversation become usable *corpora* rather than merely videoed or transcribed *data*.

3.2. Outlines for corpus development

3.2.1. Mono-modal corpus design

'Time and fiscal constraints, as well as the traditions of different research communities make it impossible to adopt a single standard for all corpora' (Strassel and Cole, 2006: 2, a fact also explored by Lapadat and Lindsay, 1999). Therefore, current mono-modal corpora, as with *developing* 4th generation MM corpora, are bespoke insofar as they are commonly designed and constructed in 'light of the investigator's goals' (Cameron, 2001: 29, also

see Lapadat and Lindsay, 1999; O'Connell and Kowal, 1999: 112; Reppen and Simpson, 2002: 93 and Roberts, 2006), in order to meet a given research need and/or to allow users to focus on specific features of spoken or written language.

Despite this, since corpus construction is generally motivated by the aim of representing an 'authentic' sample of language, the 'unambiguous, rigorous, consistent and well-documented practices [involved] in data development' (Wynne, 2005) are of a fundamental concern when designing corpora. Although such practices are to a certain extent locally determined (Conrad, 2002: 77), Sinclair offers suggestions for 'good practice' that provide general benchmarks for all corpora (2005 - see Wynne, 2005 for similar prescriptions[15]). Although these are designed with 3rd generation corpora in mind, they are also relevant for 4th generation corpora, and exist as a good starting point for discussions of MM corpus development. They are as follows:

1. The contents of a corpus should be selected without regard for the language they contain, but according to their communicative function in the community in which they arise.

2. Corpus builders should strive to make their corpus as representative as possible, of the language from which it is chosen.

3. Only those components of corpora which have been designed to be independently contrastive should be contrasted.

---

[15] Exhaustive standards for the construction of spoken corpora specifically have also been developed by EAGLES (Expert Advisory Groups on Language Engineering Standards), refer to the following website for further details: http://www.spectrum.uni-bielefeld.de/~gibbon/gibbon_handbook_1997/

4. Criteria for determining the structure of a corpus should be small in number, clearly separate from each other, and efficient as a group in delineating a corpus that is representative of the language or variety under examination.

5. Any information about a text other than the alphanumeric string of its words and punctuation should be stored separately from the plain text and merged when required in applications.

6. Samples of language for a corpus should, wherever possible, consist of entire documents or transcriptions of complete speech events, or should get as close to this target as possible. This means that samples will differ substantially in size.

7. The design and composition of a corpus should be fully documented with information about the contents and arguments in justification of the decisions taken.

8. The corpus builder should retain, as target notions, representativeness and balance. While these are not precisely definable and attainable goals, they must be used to guide the design of a corpus and the selection of its components.

9. Any control of subject matter in a corpus should be imposed by the use of external, and not internal, criteria.

10. A corpus should aim for homogeneity in its components while maintaining adequate coverage, and rogue texts should be avoided.

It is important to acknowledge that the above suggestions are theoretically idealistic. 'Since language text is a population without limits, and a corpus is

necessarily finite at any one point; a corpus, no matter how big, is not guaranteed to exemplify all the patterns of the language in roughly their normal proportions' (Sinclair, 2008: 30). Corpora are necessarily 'partial', as it is impossible to include *everything* in a corpus, since the methodological and practical processes of recording and documenting natural language are selective; ergo 'incomplete' (Thompson, 2005, see also Ochs, 1979; Kendon, 1982: 478-9 and Cameron, 2001: 71). This is true irrespective of whether a corpus is specialist or more general in nature.

Given this selectivity, the requirements for, for example, *representativeness, balance* and *homogeneity* (see suggestions 8 and 10, also see Biber, 1993) can be difficult to meticulously uphold. This problem is intensified by the fact the notions of, again, *representativeness, balance* and *homogeneity,* are relative, abstract concepts that are open to wide interpretation. A corpus that is sufficiently 'balanced' to achieve the aims of a particular corpus developer, or to allow for a specific line of research, may not be adequate for other users or lines of linguistic enquiry. Nevertheless, 'we use corpora in full awareness of their possible shortcomings' (Sinclair, 2008: 30) because there exists no better, alternative resource for the analysis of real life language-in-use than a corpus offers, nor better strategies for exploring such language than with the use of current CL methodologies.

3.2.2. A new design methodology for 4th generation corpora

Despite the potential for variety in the specific approaches used, when collecting and assembling naturally occurring qualitative data, in linguistics and beyond, there are essentially 4 fundamental processes which need to be

considered. These are outlined below (for similar models consult Psathas and Anderson, 1990; Leech et al., 1995; Lapadat and Lindsay, 1999; De Ruiter et al., 2003; Thompson, 2005 and Knight et al., 2006):

1. Recording.
2. Transcribing.
3. Coding and mark-up.
4. Applying and presenting data.

Although these processes are portrayed in a list-like format, it is appropriate to think of each as operating as part of a complete research *system*, rather than as being stages that are temporally ordered and distinct. So each stage is best conceptualised as interacting with, and influencing the next. Just *how* each of these interact, however, is reliant on the specific approaches and methods adopted as part of each stage. Again, since corpus construction is driven by the specific 'investigator's goals' (Cameron, 2001: 29), the actual methods used at each of these stages are highly variable.

Accordingly, although the following sections aim to provide a general overview of some of the *typical* conventions and strategies used for corpus construction, this is not, in any way, a definitive account of possible procedures. Instead it functions to outline *some* of the choices and challenges faced by corpus linguists developing MM corpora, in order to postulate guidelines of good practice for this. In the remainder of this chapter, these stages of recording, transcribing, coding and presentation *will* be tackled in turn, however this is simply a method of providing a coherent structure to

discussions. Consequently, the interoperability of these phases is re-addressed throughout each section.

3.3. Recording corpus data

3.3.1. Defining the 'record' phase

The *record* phase is the data collection stage. Since, as discussed in Chapter 2, few current MM corpora are publicly available, and those that are have proven to be unsuitable for exploring the line of linguistic enquiry that is the concern of this thesis, *developing* MM corpora require completely new and relevant data sets to be recorded.

It is vital that all such recordings are both 'suitable and rich enough in the information required for in-depth linguistic enquiry, and of a high enough quality' (Knight and Adolphs, 2006) to be used and re-used in a corpus database. Thus, corpus developers should strive to collect data which is as accurate and exhaustive as it can be, capturing as much information of the content and context of the discursive environment as possible (Strassel and Cole, 2006: 3, also refer back to Sinclair's suggestions in section 3.2.1). This is because the loss or omission of data cannot be easily rectified at a later date, as real-life communication can not be authentically rehearsed and replicated. Hence, it is paramount for the researcher to decide exactly what is to be recorded *a priori* to picking up a dictaphone or video camera.

This necessitates a process of planning, the importance of planning for the construction of qualitative datasets, including corpora, is discussed by Psathas and Anderson, 1990 and Thompson, 2005. Primarily, the plan helps to determine the types of *subjects* to be involved, in other words who the

participants are; how many will partake in the recordings, and so on. It also determines the *design* of the recording process, the types of data which need to be recorded; the amount; the topics that are discussed in the corpus, if specific, and how such topics are adequately covered. Furthermore, the plan helps to define the physical *conditions* under which the recordings are to take place, in other words the when and where of the recording; whether data is written, audio or visual; what equipment is used; where and how this is set up.

Often corpus developers will keep a checklist or a log of their progress throughout the construction. This not only helps to detail specific recordings, and to catalogue and organise them, but it also acts as an invaluable point of reference for discussing and/or justifying anomalies or 'gaps' that occur in the data, as well as accounting for interesting patterns that may become apparent in the subsequent analyses.

3.3.2. Blueprints for recording multi-modal corpus datasets

3.3.2.1. The recording set-up

The *conditions* used in the recording phase perhaps require the most redefinition with the onset of new MM corpus datasets. Although research using audio recordings of conversation has had a long history in corpus-based linguistics, the use of digital video records as 'data' is still fairly innovative. Granted, cameras have, in the past, sometimes been used in addition to dictaphones when collecting spoken corpora, acting as an *aide-mémoire* when compiling a corpus (see the BASE[16] corpus, for example). However,

---

[16] BASE (British Academic Spoken English Corpus) is a corpus comprised of 160 lectures and 40 seminars recorded in a variety of different academic departments at Warwick and Reading University. For more information see:
http://ahds.ac.uk/ictguides/projects/project.jsp?projectId=200

these recordings are not generally integrated into the final assembled corpus. Therefore considerations such as the quality of the recordings, the basic set-up and the type of the cameras used, and so on, took less precedence than they do with *developing* MM datasets; for which cameras are integral to the design of the record phase.

It is interesting to note that the *conditions* and *procedures* used in the VACE, AMI, MSC 1, NIST and MM4 corpora (refer to Figure 2.1 in Chapter 2 for further details and related references) are all based on a similar model; utilising a range of highly specialised equipment in a standardised, and thus *replicable*, recording set-up. This tends to be based on a variation of that seen in Figure 3.1, an example of a MM corpus recording set-up plan taken from the VACE Multimodal Meeting Corpus (Chen et al., 2005: 3).
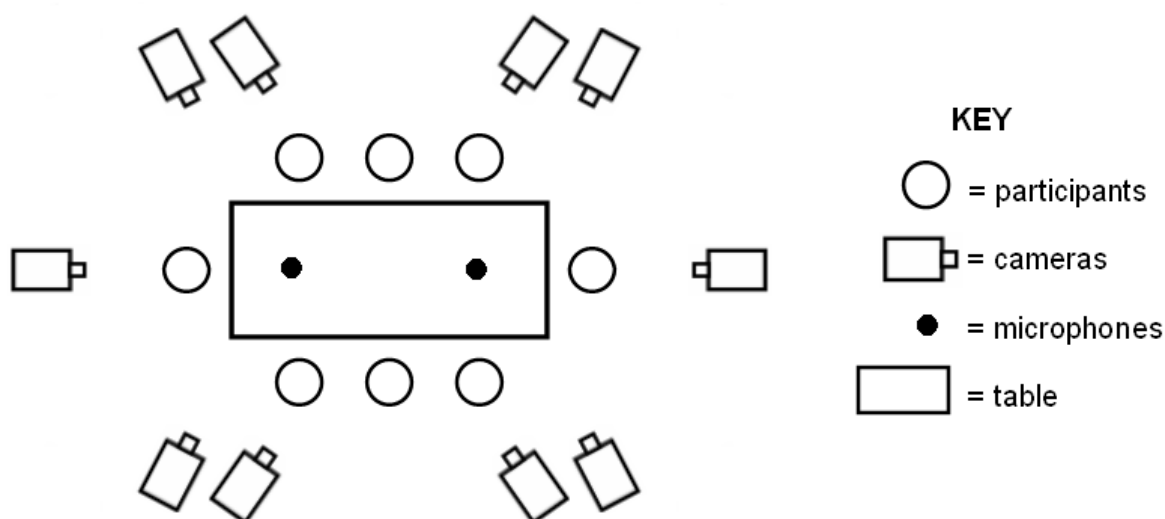
Figure 3.1: An example of the recording set-up typically used in specialist meeting room corpora (example taken from the VACE corpus, Chen et al., 2005).

The use of multiple Digital Video (DV) cameras in this set-up allows for a fairly large number of speakers (ranging from 2 to 8 in each of these corpora) to be recorded simultaneously, at a relatively close range. These cameras are either fixed on static tripods around the room, or suspended from the ceiling using overhead rail systems, as with the VACE corpus. In the case of the AMI corpus, additional remote participants are also actively involved in discussions by means of video links and conferencing software.

Each camera also records sound, which, when coupled with the output received from the fixed mounted microphones and, often, wireless microphones attached to each participant, allows for a high quality of audio output to also be collected. Each audio and video output can subsequently be synchronised, based on time, after the recording, in order to allow users to navigate the data with ease.

Given that the set-up is so fixed, it is likely that large datasets can be assembled fairly swiftly, as with the 100 hours contained within the AMI corpus, since the positioning of cameras, and so on, can be maintained from one recording session to the next. Only participants and the specific content of the discussions will change. Although, obviously this relies on the corpus compiler having the resources to, firstly, have access to this equipment and, secondly, to dedicate these cameras to corpus compilation alone, (semi) permanently fixing them into these specific positions in the recording room.

A primary criticism of the VACE corpus recording set-up, one which holds true for all forms of video recording, is that although there are no researchers or bystanders physically present throughout the recording of the data (only the recorded participants), the presence of the cameras alone can cause some of

the effects associated with the 'observer's paradox' (Labov, 1972). Participants may consciously, or even sub-consciously, adjust their behaviours because they are aware that they are being filmed, as video cameras are generally quite obtrusive. However, since it is technically not ethical to 'hide' cameras, it is difficult to minimise the potential effect that the observer's paradox will have on how *naturalistic* the participant's behaviour is.

Another shortcoming associated with this method of recording, one which perhaps limits the extent that it can be transferred beyond this specialist context, is that the fixed positioning of the table, participants and even cameras produces almost experimental, laboratory-type, conditions. Although this set-up is perhaps not strictly as experimental as that used in the MIBL corpus and the Czech audio-visual speech corpus (see Figure 2.1 in Chapter 2 for further details), it can seen to be far from naturalistic. Firstly, the use of the table means that there is a limited view of each participant, only from the torso upwards. Thus, should a researcher desire to explore, for example, leg and lower body movements or even exaggerated hand and arm movements, this would not be possible as these movements are likely to take place out of view of the camera lens. Secondly, as participants are only allowed to sit in specific locations, they are not really encouraged to, for example, get up and move around as perhaps they naturally would. This is because such movements are likely to affect the quality of recordings as they will move out of the focus of the cameras.

Since the cameras that are used are static, the data collected is very much fixed in terms of location and time. This set-up does not support recordings of spontaneous interaction in real-life environments 'on-the-move'. It is relevant

to note that, as discussed in Section 2.2.4 of Chapter 2, both the SVC corpus and the SK-P 2.0 corpus (see Figure 2.1, Chapter 2) begin to tackle this limitation by utilising a corpus recording approach which is less context-specific, thus more 'mobile'. The SVC, for example, uses portable Smartphone devices to record a range of different public spaces, some:

Indoors (office, lobby, public cafe) and some outdoors (courtyard, park) with varying acoustic and lighting conditions, changing sources of background noise and visual background (resulting for example from different weather conditions: sunny with blue sky or cloudy). These conditions were not controlled for the experiment but have been documented in the recording protocol. (Schiel and Mögele, 2008: 2)

Similar environments were recorded as part of the SK-P 2.0 (see Figure 2.1, Chapter 2 for further details, also see Schiel et al., 2002).

In theory, this variability starts to overcome some of the drawbacks of using laboratory-type settings for recording MM corpora. However, in reality these corpora do not exist without shortcomings of their own. Primarily, the Smartphone devices are only used to record single participants in these corpora, even despite the fact the SVC is based on dyadic conversations. This limits the potential for exploring patterns in dyadic or group behaviour in the data. Furthermore, the quality of these recordings is not particularly good and only specific sequences of behaviour, facial expressions and, in this case head movements, can be captured at a high resolution. However it is

appropriate to note that this is perhaps more a limitation of the equipment than the recording design methodology. An additional, more general limitation of these corpora is that they are both task-orientated, so although discourse is occurring in natural contexts, the prescribed nature of the tasks involved affects the spontaneity and perceived *naturalness* of the data collected.

Despite this, these corpora can be seen to offer an insight into possible directions that linguistic corpora development may take in the future; an insight into the type of corpus datasets that will possibly *supersede* 4th generation MM corpora. Indeed plans for similar 'mobile' corpus datasets, comprising ubiquitous information are being drawn-up by researchers at the University of Nottingham, as part of the DReSS II[17] project. This includes data from a range of different contexts, including face-to-face situated discourse through to the use of SMS messages, MMS messages, and interaction in virtual environments and so on. The DReSS II project aims to utilise digital technologies to develop a system for recording the language experience of individuals from multiple perspectives. This is with the view of enabling a more detailed investigation of the interface between various different communicative modes; tracking a specific person's (inter)actions over time, i.e. across an hour, day or even week. The analysis of information of this kind can potentially help to question the extent of language choices determined by different communicative environments. Such advances will help to overcome some of the limitations of current MM corpora, i.e. those associated with context-specificity; the observer's paradox; fixed and static recording method, the perceived 'naturalness' of data, and so on. Furthermore, they will perhaps

---

[17] More information on DReSS II can be found at:
http://www.ncess.ac.uk/research/digital_records/

allow us to gain a better insight of true, 'real-life' language-in-use as indeed corpora aim to provide (refer back to Sinclair's suggestions, 2005).

Studies into corpora of this nature are therefore very much a priority for the future in CL research and development. However, at present no fully functioning corpus of this nature is in existence because linguists are *still* tackling the problems associated with MM corpora of the nature as discussed in the current thesis.

3.3.2.2. *The recording set-up used for the NMMC*

Again, the NMMC, as with the CID, IFADV and the Göteborg Spoken Language Corpus (refer to Section 2.2.4 of Chapter 2 for more details) was designed to allow more flexibility in the recording of natural language data than, more experimental, specialist meeting room corpora such as the VACE corpus allow. This was in order to meet the following prescriptions (Knight 2006, in alignment with Sinclair's prescriptions, 2005):

- To record *multiple modes* of communication in *natural contexts.*
- To use a recording method that can be *easily replicated* in future studies.
- To record both the *individual* sequences of body movements of all speakers in an interaction, but allow for the analysis of *synchronised* videos in order to allow the examination of co-ordinated movement (i.e. across each speaker).
- To obtain recordings that can be *replayed* and *annotated* by other researchers.

However, as with the corpora noted above, it proved difficult to strike a balance between the resolutions of recordings, i.e. the *quality* of data collected, and the perceived *naturalness* that it represents. Furthermore, it was even more difficult to maintain a balance between these factors and the usability of the corpus data collected. Consequently, the basic recording set-up used for the NMMC is thus somewhat still similar to the laboratory-type settings seen with the VACE corpus, and other corpora listed above. However this was not merely restricted to a meeting room environment. Figure 3.2 presents a plan of this set-up (Knight et al., 2009).
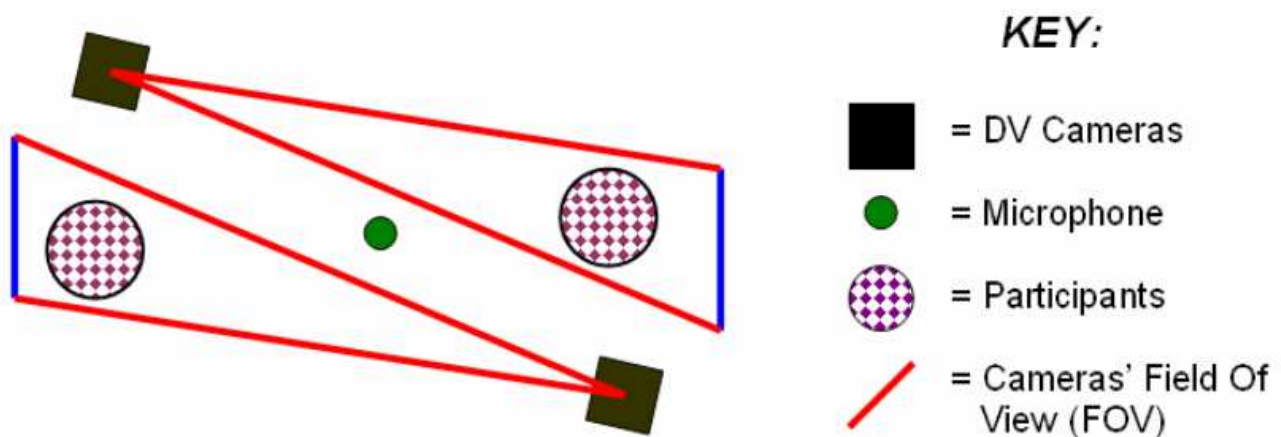
Figure 3.2*:* A basic recording set-up for multi-modal corpus development, based on the NMMC.

Two DV cameras were used as part of this set-up, specifically to allow for *individual* bodily movements of each participant to be recorded and also enabling the data to be digitised for subsequent Mpeg compression. These

images were later synchronised using Adobe Premiere[18], so that the behaviours of both participants could be observed simultaneously during the analysis of the data.

These recordings took place in relaxed, familiar settings' with 'each conversation last[ing] 45-60 minutes (see Knight et al., 2006). The purpose of this was 'to minimise the effects of observer's paradox, by enabling speakers to become more at ease around recording equipment, thus promoting talk that is as natural as possible. Although the setting used was perhaps more laboratory-like than 'natural', as Argyle notes, it *is* actually possible to arouse innate responses and patterns of behaviour from participants in such environments (1988: 11), provided that they feel relaxed and at ease with, amongst other things, the settings and the people with whom they are communicating.

To enhance the quality of audio data collected, a high specification microphone was positioned between speakers. For the purpose of recording the CID, this microphone was supplemented by head-set microphones for each participant. This was to allow the corpus to be utilised for the explorations of the phonetic characteristics of talk, which is one of the key aims of the CID. Similar devices were not used in the NMMC as it was decided that the addition of such headsets would likely to obscure the images of the head, face and upper torso, making it difficult to explore specific sequences of movement in such areas with ease, as is the concern of the present study.

---

[18] Adobe Premiere Pro is sophisticated digital editing software developed by Adobe. For more information see: http://www.adobe.com/products/premiere/

Further to this, unlike the set-up seen in Figure 3.1, participants were not specifically requested to sit around tables for the NMMC. This was to enable recordings to capture a range of different forms of NVB and NVC, focusing not only on the head and face, but on the hand and arm movements, and the complete torso of each participant. This was to enable a range of different iconic gestures and certain proxemic movements to be studied.

Although the conversations recorded for the NMMC were not strictly task-driven it is important to note that all data was collected from a university setting. All episodes featured native English speakers in academic environments at the University of Nottingham. These conditions perhaps suggest that the results from any analyses of such data are likely to be somewhat context and/or genre dependent. Although this is obviously a shortcoming of the corpus, perhaps aligning it to a more 'specialised' type, this restricted cross section of participants exists here as a useful starting point for the development and analysis of new MM methodologies. However, it would be beneficial if data from a wider range of socio-cultural contexts collected under different conditions were available for future MM CL research.

3.3.2.3. Corpus size

The question of *how much data is enough?* when constructing a MM corpus is a complex and challenging one, for which no definitive answer exists. This is true not only for MM corpora, but is also relevant for mono-modal corpora. On the topic of corpus size, Baroni and Ueyama (2006) suggest that:

Because of Zipfian properties of language, even a large corpus such as the BNC contains a sizeable number of examples only for a relatively limited number of frequent words, with most words of English occurring once or not occurring at all. The problem of 'data sparseness' is of course even bigger for word combinations and constructions.

In 1935 Zipf used a counting based method to ascertain the frequencies of various linguistic features in order to extract interesting observations in respect of real-life language use. As a result of his pioneering work, 'Zipf's law' (1935) was proposed, suggesting that 'the product of rank order and frequency [of lexemes] is constant' (Kilgarriff, 1996: 39) in language. So, in theory, this implies that 'the most common word in a corpus is a hundred times as common as the hundredth most common, a thousand times as common as the thousand, and a million times as common as the millionth' (Kilgariff, 1996: 39).

This constant suggests that a key 'factor that affects how many different encounters you have to record [for a corpus] is how frequently the variable you are interested in occurs in talk' (Cameron, 2001: 28). Thus, larger datasets, or indeed datasets from specific contexts, will be required for less common words, whereas with more commonplace phenomena this is not always necessary.

So there is little point in collecting, for example, 70 hours of video data to explore the presence of *yeah* in discourse when the results would probably

not be any more revealing than those seen in 7 hours of data, given that this minimal response is so frequent (see Beach, 1993; Drummond and Hopper, 1993a and Gardner, 2001). Whereas, if 70 hours of data only includes a couple of instances of the phenomenon under focus it is prudent to think of other ways of collecting relevant data, or indeed to reconsider whether it is more cost-effective to focus upon something that is more frequent in discourse.

Referring to spoken corpora specifically, Thompson (2005) highlights the necessity of deciding between the 'breadth' and 'depth' of what is to be recorded, and for providing a cost-benefit analysis of this. This notion of the cost-benefit is also relevant for emergent MM corpora. Essentially, this identifies the relative advantages between capturing large quantities of data, in terms of time and the number of encounters or discourse contexts recorded, the amount of detail in which is annotated and analysed, and the extent to which this optimizes the quality of results obtained following such analyses.

Theoretically, a non-specialised MM corpus, i.e. one which is built for general purposes rather than to answer a specific research question, should perhaps aim to provide (i.e. contain) data which includes a range of different speakers in a range of different discursive contexts. This would include participants of different ages and genders from a variety of socio-cultural backgrounds speaking in different conditions, from monologic talk to dyadic and group scenarios. However, in reality it would require much time and many resources to collect such data, so in terms of practically it is unlikely that this can ever be fully achieved.

Despite this, it remains relevant to suppose that although this notion of the optimum size of a corpus is difficult to quantify, regardless of whether mono-modal or MM, larger size datasets, perhaps comparable to the 100-million-word BNC, will best counter Chomskyian criticisms of the unaccountability of *small* corpora (see Chapter 2 for further details). Given that the technological and methodological procedures used in MM corpus construction and analysis are still *developing*, multi-million-word MM corpora have yet to be realised, although it is hoped that they will be available in the future.

3.3.2.4. Metadata for multi-modal corpora

Apart from recording the actual episodes of interaction between speakers, an inherent part of corpus development involves the construction of records of *data about data,* i.e. 'metadata'. Metadata is critical to a corpus as 'without metadata the investigator has nothing but disconnected words of unknowable provenance or authenticity' (Burnard, 2005).

Again, since corpora are inherently selective, the addition of metadata archives the key facets of this selectivity; detailing the recording techniques and equipment used, the speakers involved, and the context(s) of the interaction. Reference to these factors can assist in understanding patterns that emerge when analysing communicative datasets, and can help to start to re-contextualise and account for some of the behaviours seen. Metadata information is commonly integrated into corpora as part of the coding and annotation process (as discussed in Section 3.5, below) and, as with other elements of coding, there are various different ways in which aspects of metadata are physically annotated in corpora.

While there are no universal prescriptions for defining which features are marked up as part of corpus metadata, how this is structured or how it becomes searchable within the database, Burnard (2005) suggests that it is essential to include details of the editorial, analytic, descriptive and administrative processes of corpora composition. These categories have been used to annotate the BNC[19]. The inclusion of this information assists in identifying the name of the corpus (administrative metadata), who constructed it, and where and when this was undertaken (editorial metadata) together with details of how components of the corpus have been tagged, classified (descriptive metadata), encoded and analysed (analytic metadata).

Burnard's categories provide a suitable benchmark for metadata description in MM datasets since the large majority of the elements discussed as part of the corpus development methodology in this chapter, are qualified, in some way, using these four categories. However, it is important to emphasise that this issue of metadata description, classification and codification requires further discussion and revision as MM corpora become more large-scale and mainstream in corpus research and linguistic methodology.

3.3.2.5. A note on ethics

There are many ethical concerns requiring consideration as part of a MM corpus design methodology. These fall into the following broad categories:

---

[19] For further information on the conventions for encoding the BNC please see:
http://natcorp.ox.ac.uk

- *Institutional*: Guidelines prescribed by a particular University (imposed by a central Ethics Committee) and/or department.

- *Professional*: Common guidelines used across a specific discipline, research paradigm and/or research funding council[20].

- *Personal*: Personal and/or collaborator defined ethical standards which exist to maintain relationships and integrity in research.

Moral and legal obligations faced at each of these levels can heavily influence processes undertaken during every stage of the corpus development; from the data collection phase through to its presentation and analysis.

Current practice (i.e. ethics on a *professional* and/or *institutional* level) suggests that corpus developers should ensure that formal written or video and/or recorded consent is received from all participants involved; *a priori* to recordings. Conventionally, this consent stipulates how recordings will take place, how data will be presented and how/for what research purposes it will be used (Leech et al., 1995 and Thompson, 2005). While a participant's *consent to record* is relatively easy to obtain, insofar as this commonly involves a signature on a consent form, it is important to ensure that this consent holds true for *every stage* of the corpus compilation process.

It is also appropriate to receive *consent to distribute* recorded material, because although a participant may be happy to record a conversation they may not be as willing to freely offer this consent if they know how the data will be used. This is especially true if the data is to be published and distributed

---

[20] For example, see the ethical guidelines provided by the ESRC, Economic and Social Research Council: http://www.york.ac.uk/res/ref/ Also see the 'Recommendations for Good Practice in Applied Linguistics', provided by BAAL, the British Association for Applied Linguistics: http://www.baal.org.uk/about_goodpractice_full.pdf

widely, or if it is to be used in environments where an individual's peers are present.

Paradoxically, in reality it is difficult to determine to what extent consent can be truly informed. This problem commonly exists as an ethical concern on a more *personal* level. Although exhaustive descriptions of specific processes of recording and/or constructing a corpus are provided to participants, unless they themselves are perhaps a corpus linguist, and are familiar with procedures, or a researcher accustomed to CL methodology, participants may still not fully understand to what they are contributing. So, although they technically provide 'informed consent', the validity of this status as being 'informed' can be questioned.

A further ethical concern involves the notion of *anonymity* in data. Traditional approaches to corpus development emphasise the importance of striving for anonymity when developing records of discourse situations, as a means of protecting the identities of those involved. To achieve this, the names of participants and third parties are often modified or completely omitted, along with any other details which can make the identity of participants obvious (see Du Bois et al., 1992; 1993). The quest for anonymity can also extend to specific words or phrases used as well as topics of discussion or particular opinions deemed 'sensitive' or 'in any way compromising to the subject' (Wray et al., 1998: 10-11).

Anonymity is relatively easy to address when constructing written-based, mono-modal corpora. In such cases, if the data used is already in the public domain and freely available, no alterations to the texts included are usually required. If not, permission needs to be obtained from the particular authors or

publishers of texts, i.e. its copyright holders, and specific guidelines concerning anonymity can subsequently be discussed and addressed with these authors, with alterations to the data made as necessary.

Similar procedures are involved when constructing spoken corpora. Since these corpora are generally presented in text-based formats, modifications, omissions and other such measures of anonymity can be undertaken at the transcription phase of corpus development. This allows participants who have already provided their consent to be involved in the process.

Anonymity is more problematic when physically integrating the actual audio records of conversations into the corpus database. Audio data is 'raw' data which exists as an 'audio fingerprint' insofar as it is specific to an individual. This makes it relatively easy to identify participants when audio files are replayed. Therefore, it is logical to suggest that to achieve anonymity in audio files, the nature of the vocal input should be altered in some way in order to make the participant less recognisable. However, to allow the files to be adequately used for, for example, the exploration of phonetic patterns associated with particular word usage, any such alteration or 'tampering' with the audio streams can result in data that is misleading or misrepresentative. Undoubtedly, it would be possible to protect the identity of speakers using actor's voices, although this procedure would again forfeit the authenticity of the data, by compromising the spontaneity and 'naturalness' of the talk. Regardless of how accomplished the actor is, it unlikely that every acoustic or prosodic feature can be adequately recreated.

A similar problem concerning anonymity is faced with the use of video data. Although it is possible to shadow, blur or pixellate video data, in order to

conceal the identity of speakers (see Newton et al., 2005 for a method for pixellating video), these measures are difficult to accomplish especially when dealing with large datasets. In addition, such measures obscure the facial features of the individual, blurring distinctions between gestures and language forms. This again results in datasets that are unusable for certain lines of linguistic enquiry. If, for example, the researcher desires to use the corpus to explore facial expressions or eye movements, or even head nods, as is the concern of the present study, pixellisation would inhibit their ability to do so.

Before going to such lengths in the quest for anonymity in data, it is perhaps relevant to question whether it is necessary to consider anonymity in such a controlled way at all. If participants have provided written permission to be recorded, they are in effect providing consent for their image and/or voice to be used, since people themselves are not anonymous. In short, it may be nonsensical to conceal these features when creating a database of *real-life* interaction as by altering or omitting the participant's identities, the data becomes far from *real*. The matter of protecting the identity of third parties, however, remains an ethical challenge with such data, along with the issue of re-using and sharing contextually sensitive data recorded as part of MM corpora.

In sum, the corpus developer is required to strike a balance between the quest for anonymity in the data and its usability and accuracy for research; a balance that is appears difficult to achieve. However, it is valid to note that if, for example, a corpus is intended for small-scale studies and is to be used only by those involved in its development, the requirements for anonymity are

unlikely to be as complex or stringent as for large-scale corpora that are intended for future general release.

Given that the present study and the NMMC, in general, only uses a small amount of data from a limited cross-section of participants, these problems of 'ethics' are perhaps not particularly relevant or complex here. Each participant signed *permission to record* forms, and provided consent for conversations to be analysed as part of this thesis. If the content used were to be widely published, the question of ethics would need to be re-addressed, although at present the likely small readership of this study means that this is not the case.

In terms of future large-scale MM corpus development, it is important to reconsider these ethical requirements and attempt to draw up some new guidelines and appropriate procedural blueprints for dealing with MM data, in order to adequately protect participants and developers from ethical or legal problems which may arise. In short, regardless of the strategies used, it is paramount that there is a consistency between these measures, across all three modes of data (i.e. the textual, spoken and visual), as it would be counter-productive to exhaustively omit or alter details in the written transcript when the corresponding audio files remain unchanged, and vice versa.

3.4. Transcribing corpora

3.4.1. Current transcription methods

The second phase of the MM corpus development methodology, transcription, is seen as 'an integral process in the qualitative analysis of language data' one which is widely employed in applied research across a number of

disciplines and in professional practice fields' (Lapadat and Lindsay, 1999: 64). Transcription is commonly conceptualised as a type of research method, a 'process reflecting theoretical goals and definitions' (Ochs, 1979: 44 also see Edwards, 1993 and Thompson, 2005).

Ochs (1979: 44) suggests that it is at the point of transcription that spoken words technically become language data; when it becomes a document of a written or graphic format that represents something else. So it instead becomes an abstract, physical manifestation of that vocal stimulus (see Cameron, 2001: 73). Accordingly, a transcript is often viewed as being 'both interpretative and constructive' (Lapadat and Lindsay, 1999: 77, also see O'Connell and Kowal, 1999: 104); providing a window into communicative events from the perspective imposed by the person(s) responsible for the transcription.

As with all stages of corpus construction, 'there is little agreement among researchers about standardisation of [transcription] conventions' (Lapadat and Lindsay, 1999: 65). No strictly 'standard' approach is used to transcribe talk in CL research (Cameron, 2001: 43).

Efforts have been made to standardise transcription *beyond* the specific scope of CL methodology. Gail Jefferson's Transcription System (Jefferson, 2004), based on CA methodologies (see Markee, 2000 and ten Have, 2007), outlines some shared conventions of transcription for use in linguistic research. This system is now widely used by conversation analysts and a host of other researchers working with language data (see Psathas and Anderson, 1990: 75). However, although the Jefferson coding scheme is sufficient in

meeting the needs of CA researchers directly, it is not fully transferable to CL based methodologies.

Given this, Leech et al. acknowledge that the 'need to converge towards a standard, or (to weaken the concept) towards agreed guidelines is becoming a matter or urgency' (1995: 5) in CL methodology. Such an agreement, a consistency in transcriptions conventions, would theoretically allow data to be transferable and re-usable across individual corpus databases. At present 're-use is a rare phenomenon' in language research, (Dybkjær and Ole Bernsen, 2004: 6). Although there are naturally many ethical challenges associated with this, it would essentially allow both the size and quality of corpus data available for linguistic research to be enhanced, without individuals or specific teams of researchers expending large amounts of time and resources.

3.4.2. Transcribing multi-modal corpora

3.4.2.1. Methodological considerations

The key question that needs to be addressed when transcribing MM datasets is how, if at all, characteristics of speech and gesture-in-talk are to be documented in the textual record that is presented in the corpus interface; i.e. should one attempt to textually mark-up visual, and concurrent verbalised features in the transcript, or should such features be kept distinct?

Commenting on transcribing spoken language, Schiffrin suggests that the use of a 'transcription system that builds on graphic punctuation symbols forces us to think of such chunks as sentences, rather than as providing an accurate presentation of how speakers themselves produce language' (1994: 25). Thus, by transcribing audio stimuli we are effectively losing some of the

'truth' of the language production as the reduction of speech into lexical forms, i.e. the use of graphic representations, as presented in the transcript, cannot wholehearted depict all aspects of talk. Whether this is *at all* possible is another question, however.

This limitation is intensified when attempting to answer the question of whether, and how we should transcribe forms of gesticulation as 'when transcribing gestures, especially in manual annotation, a lot of information is lost compared to the complexity of the original movement' (Kipp et al., 2007: 325). Again the 'difficulty of fluidity' means that unlike words, gestures are not readily made 'textual units' (Gu, 2006: 130) so have no standard text-based methods for their representation in transcript form. So, while linguists are familiar with attempting to transcribe speech, even if this is only a partial representation of the truth, the transcription of gesticulation is less prescribed and more difficult to embark on.

Having said this, while, at present, concordancers and CL software use lexis as the only entrance point to data searches, the addition of multiple *forms* of representation *beyond the text*, means that MM corpora are not necessarily restricted to this. So this problem of transcribing the untranscribable, i.e. converting forms of gesture into textual units, is perhaps no longer strictly applicable. In other words, one method of solving the challenge of transcribing the MM may perhaps be to simply restrict, as a 'reference point', the exploration of gestures to the visual medium rather than attempting to include references of these phenomena within the textual transcriptions. This method would instead mean that the process of quantification through textual representations is completely avoided. In this

case the researcher would instead be required to search for specific sequences of gesticulation either manually, by replaying given video sequences, or through some form of automated technique (presuming the video data has been pre-coded for exploration, see Section 3.5); in both cases, examining the video data alone. These features of interest can then be extracted and analysed in conjunction with the transcribed spoken words where required.

However, given that manual searches of data are arduous and automatic searches are not completely reliable at present, insofar as no 100% accurate real-time movement tracker and coding tool is in existence (see Chapter 5 for further details), such a method is far from practical. Therefore logistically speaking, some form of annotation and mark-up of visual data is currently necessary to facilitate the analysis of MM data (further details of coding and mark-up are discussed in Section 3.5). This may comprise specific annotations which are integrated directly into a transcript of speech, or may consist of an entirely different movement-focused textual transcript or, finally, exists as a separate coding track which is time-aligned with the speech-based transcript.

Regardless of the method used, it is important, for the future of MM corpus research development, that a more integrated and standardised system for MM transcription is compiled, a system which incorporates 'criteria that show how different resources contextualise each other' (Baldry and Thibault, 2001: 88), helping to effect 'a transition from *MM transcription* to *MM corpus*' (2001: 90). Such conventionalised integrated frameworks have yet to be devised.

3.4.2.2. Tools for transcription

A wide range of computer software exists that enables researchers to transcribe audio and/or video records of communication digitally. Transtool, Tractor, TraSA and SyncTool, for example (refer to Allwood et al., 2001, for more information on each of these), have provisions for transcription; annotation; coding scheme creation (a matter discussed more fully in Section 3.6) and/or for visually integrating different modes of data for subsequent analysis. Other tools such as MultiTool[21]; Transcriber; iTranscriber[22] (both used in the VACE corpus, see Figure 2.1) and MuTra[23] (used in the MIBL corpus) assimilate these features, allowing the researcher to 'simultaneously display the video and relative orthographic transcription of dialogues so that the operator can easily observe when gestures are produced together with speech' (Cerrato and Skhiri, 2003: 255, discussing MultiTool specifically).

Transana[24] has similar functionalities to these tools. Not only does it provide the means for researchers to transcribe and edit their own datasets, it enables the alignment of transcriptions with video and/or audio records through the use of a time-stamping facility. An example of a time-stamped transcript excerpt, completed using Transana, can be seen in Figure 3.3.

---

[21] MultiTool is a multimodal transcription and analysis tool, freely available from: www.ling.gu.se/projekt/tal/multitool/

[22] More information about Transcriber and iTranscriber is available online, although the tool is not freely available for download: http://www.icsi.berkeley.edu/Speech/mr/mtgrcdr.html

[23] MuTra is a freely available multimodal transcription tool available from: www.swrtec.de/swrtec/mibl/mutra/

[24] Transana is qualitative analysis software for video and audio data, developed by the University of Wisconsin-Madison Centre for Education Research. See: www.transana.org/

```
¤<0><$2> to introduce to you. She's got a fascinating title for us this afternoon
we've got on the Reindeer road with the Frodi's meat haired woman+

¤<8111><$1> Meal.

¤<9291><$2> +Sorry Meal haired woman. Okay I shall hand over to+

¤<13881><$1> Thank you <$2>.

¤<14265><$2> +Thank you.

¤<15189><$1> Um right er deepest apologies if you were expecting lepers and
lunatics I'm really sorry I can't do lepers+

¤<24766><$2> <$E> Laughter <\$E>
```

Figure 3.3: An excerpt of a time-stamped transcript, taken from the NMMC.

The timestamps, starting from 0 milliseconds, provide reference points on which to 'hang' time-series data together, such as video and/or audio files, aligning them with similar time-based records across the different data streams in the corpus. So when specific turns are highlighted in the transcript, the video and/or audio records jump to the instances where these turns are uttered. This time-stamping allows the different modes of data to be navigated systematically and with ease, making it invaluable as a point-of-entry for the analyses of MM datasets. For this reason the NMMC, and the data used in this thesis, was transcribed using Transana (additional reasons for choosing this tool are explored in Brundell and Knight, 2005).

It is important to note that time-stamps were administered on a turn-by-turn basis for the NMMC. When attempting to represent, for example, overlaps and interruptions in talk (which are commonplace in spoken communication, refer to Sacks et al., 1974), when using this approach, the analyst is required to temporally order one turn before the next as the time-

stamping facility does not allow the input of two episodes simultaneously. This method is, therefore, possibly open to question as in reality speech is rarely so 'orderly' and people do not necessarily interact in such a regimented way. Regardless of whether one simultaneous contribution is positioned only a few milliseconds before or after the other, this basic method of ordering turns in the transcription perhaps gives discourse a structure it does not, in reality, possess (Graddol et al., 1994: 182). This criticism is particularly relevant if, for example, four or five speakers are present in the conversation.

Given this, it may be more appropriate to provide distinct time-stamped transcripts for each speaker in a conversation, each of which can be individually time-stamped and aligned within the corpus interface (see Section 3.6 for further details). Alternatively, it may be appropriate to attempt to time-stamp on a word-by-word-scale rather than by turns. In fact, as Graddol et al. discuss, in reference to the general representation of speech in transcription, 'any number of complex layouts could, in principle, be devised in an attempt to provide a more valid account of interactions, although there will always be something of a tension between validity and ease of reading' (1994: 185). So, before administering such techniques it is necessary to assess the cost-benefit of using such methods; assessing what these actually *add* to the analyses and whether they are actually really required, given the amount of time and effort that they are likely to take to assemble.

Currently, word-by-word time-stamping cannot be undertaken automatically with any real degree of accuracy. It is also difficult to do this manually, since each single word needs to be assigned a time code in turn. This means that it is unlikely that large quantities of data can be processed in

such a way, with either speed or ease. Given this, and given the fact that the current thesis only deals with dyadic conversations rather than group environments (which are likely to be rife in overlaps etc), the basic methods for time-stamping and transcription, seen in Figure 3.3, are used throughout. This comprises turn-by-turn time-stamping, with both speakers from each dyadic conversation included in the same, single, transcript.

3.4.2.3. Transcription methods used in the NMMC

For the purposes of continuity, the audio recordings included in the NMCC have been transcribed by highly trained linguists adopting the same conventions used in the CANCODE corpus (see Adolphs, 2006: 134-135). As a measure of quality control, all transcripts were checked and double checked during this transcription phase. This helped to ensure that there was consistency between, for example, the orthographic representation of common but non-standardised vocalisations, which may be spelt in a variety of different ways (such as the lexemes *mmm, mm, mmmm* and *mhm*). When constructing MM corpora it was necessary to define and distinguish between such terms early on, in order to establish standardised lexical forms for their representation. This assists in ensuring that accurate and reliable analyses of the data can be conducted in the future.

The CANCODE conventions are designed to present conversational data 'in a way that is faithful to the spontaneity and informality of the talk, but is also easily accessible to readers not familiar with the conversational literature or phonological/ prosodic features' (a key requirement of transcription, outlined by Eggins and Slade, 1997:1-2). This means that, for example,

annotations of prosodic or phonetic features of talk using the IPA (International Phonetic Alphabet, see Laver, 1994 and Canepari, 2005) are not integrated within these transcripts. This is because such information would make the corpora inaccessible to researchers inexperienced in dealing with the IPA, as IPA based transcripts are both difficult to read and too specific in focus for such users. Obviously, if the corpus was intended to be of a more specific nature, with a primary function of allowing phonetic research, IPA transcription would be required. However, the relative cost effectiveness of this needs, again, to be determined *a priori* to transcription since IPA based transcription is also very time consuming.

A key advantage of MM corpora is that the actual audio files of conversations are presented to the user *in addition* to the transcription of talk. So even if the IPA is not used to annotate speech in the transcript, the integration of the audio records, possibly comprising separate audio tracks or audio derived from a video file, means that phonetic enquiries can be addressed in real-time with direct reference to these records.

3.5. Coding and marking-up corpora

3.5.1. Coding conventions

Coding is the next phase of the corpus development process. This stage involves 'the assignment of events to stipulated symbolic categories' (Bird and Liberman, 2001: 26, also see Brundell and Knight, 2005). This is where qualitative records of events start to become quantifiable, as specific 'items relevant to the variables being quantified' are marked up for future analyses (Scholfield, 1995: 46). Coding is closely linked to the transcription phase,

however, instead of providing written accounts abstracted from spoken interaction, it provides abstract definitions of these abstractions.

Coding and annotation is commonly undertaken with the use of computational software. Some current corpora are described as being un-annotated, without tags and mark-up, although the majority *are* annotated, because the addition of such annotations allow corpora to be navigated using digital software.

These annotations exist from a word-based level (*tagging*) through to a more sentence- and text-based level; involving 'the addition of typographic or structural information to a document' (*mark-up*; see Bird and Liberman, 2001: 26). Corpora can also be annotated at a higher, discourse-based, level wherein specific semantic or pragmatic, function-based codes are added. In short, various features of the discourse can thus be annotated, such as information on speakers (demographic), contextual (extra-linguistic information), P-O-S (part of speech- a form of grammatical tagging, such as the CLAWS[25] 'word class annotation scheme' used in the BNC, see Garside, 1987), prosodic (marking stress in spoken corpora), phonetic (marking speech sounds) features, or a combination of these (for more information see Leech, 2005 and McEnery and Xiao, 2004, also refer back to the metadata section in 3.3.2.4).

Early standards for the mark-up of corpora, known as the SGML (Standard Generalised Mark-up Language), have generally been succeeded by XML, (Extensible Markup Language, see Ide, 1998). These standards were developed in the 1980s when the electronic-corpora 'revolution' was just

---

[25] CLAWS, the Constituent-Likelihood Automatic Word-Tagging System, is a system for tagging English language texts (according to P-O-S). For more information see: http://ucrel.lancs.ac.uk/claws/

beginning to take off, with the transition from 1st to 2nd generation corpora (refer back to Section 2.2.2 of Chapter 2 for further details). SGML was traditionally used for marking up features such as line breaks and paragraph boundaries, typeface and page layout; providing standards for structuring both transcription and annotation.

Modern advances in technology, and associated advances in the sophistication of corpora and corpus tools, have prompted a movement towards a redefinition of SGML. Since the late 1990s, efforts have been made to establish some 'encoding conventions for linguistic corpora designed to be optimally suited for use in language engineering and to serve as a widely accepted set of encoding standards for corpus-based work' (Ide, 1998: 1, discussing the Corpus Encoding Standard, CES[26], specifically). There are various schemes of this nature, including the Open Language Archives Community (OLAC[27], see Bird and Simons, 2000); the CES; the ISLE[28] Metadata Initiative (IMDI, see Wittenberg et al., 2000) and the TEI[29] (Text Encoding Initiative, as used in the BNC, see Sperberg-McQueen and Burnard, 1999).

In general these schemes aim to cater for corpora of any size and/or form, including spoken and/or written corpora, specialised and/or general corpora. Thus, they work on the premise that the standardised nature of corpus encoding conventions will allow coded data and related analyses to be re-used and transferred across different corpora. However, while many of these

---

[26] More information about the CES can be found at: http://www.cs.vassar.edu/CES/
[27] OLAC aimed to provide a 'common framework across electronic preprint archives'. For more information see: http://www.language-archives.org/docs/white-paper.html
[28] Details of the ISLE project can be found at the following website: http://isle.nis.sdu.dk/
[29] The TEI is 'a consortium which collectively develops and maintains a standard for the representation of texts in digital form'. For more information on the TEI see: http://www.tei-c.org/index.xml

schemes share some similarities, and the same intentions in respect of standardisation, at present there remains no universally-used prescribed method of corpus mark-up and encoding, although TEI is perhaps currently the closest to this.

As with the *record* and *transcription* phases, the level of detail used in the coding phase, 'the actual symbolic presentations used' (Leech, 2005) when annotating a corpus, is thus generally dependent on the purpose and aims of the corpus (i.e. they are 'hypothesis-driven', refer to Rayson, 2003: 1, also see Allwood et al., 2007a). So, 'there is no purely objective, mechanistic way of deciding what label or labels should be applied to a given linguistic phenomenon' (Leech, 1997: 2). However, it should be noted that regardless of the standards and systems of notation used to encode corpora, the majority tend to integrate this information into the corpus in the same way. Specific codes and tag-sets are usually integrated within the underlying infrastructure of a corpus, contained within searchable header information, separating the 'extra-textual and textual information' from the 'corpus data (or transcripts) proper' (McEnery et al., 2006: 23). This is usually XML based.

3.5.2. Gestural coding schemes

It is important to note that while the majority of current encoding schemes and approaches deal with the mark-up of selected extra-linguistic information, they do not have provision for marking up discourse *beyond the text* in any great detail, insofar as they are not fully extendable to all MM features of talk. As Baldry and Thibault indicate (2006: 148):

In spite of the important advances made in the past 30 or so years in the development of linguistic corpora and related techniques of analysis, a central and unexamined theoretical problem remains, namely that the methods adapted for collecting and coding texts isolate the linguistic semiotic from the other semiotic modalities with which language interacts.... [In] other words, linguistic corpora as so far conceived remains intra-semiotic in orientation.... [In] contrast multi-modal corpora are, by definition, inter-semiotic in their analytical procedures and theoretical orientations.

Thus within the field of linguistics, no scheme really exists with the capacity to fully support the mark-up of NVC or NVB, nor do they integrate information from both spoken and non-verbal stimuli. However, there are many schemes which deal with the coding and annotation of visual and/or multi-modal datasets, and associated methodological approaches to the application of these, beyond the area of AL (Applied Linguistics) and CL research. Therefore, it is relevant to discuss these briefly here.

Firstly, there are a wide variety of coding schemes which concentrate solely on facilitating the mark-up and labelling of gestures according to kinesic properties. These function to explicitly define the specific action, size, shape and relative position of *movements* that comprise forms of gesticulation (see Frey et al., 1983; McNeill, 1992 and Holler and Beattie, 2002, 2003, 2004 for examples of these). One widely used scheme of this nature, the FACS coding scheme (Ekman and Friesen, 1978- for examples of studies that use FACS see Buck 1990; Black and Yacoob, 1998; Pantic and Rothkrantz, 1999; 2000;

Kanad et al., 2000; Tian et al., 2000; Kawato and Ohya, 2000 and Rosenberg et al., 2001) is perhaps the one which is most relevant to the current thesis, in that it specifically deals with head movements (in addition to facial expressions).

FACS provides the referential guidelines for appropriately sub-dividing and encoding a facial image generated from a video recording, according to key 'motion reference points', defined by specific facial muscles known as Action Units (AUs, see Ekman and Friesen, 1978). There are 46 different locations of AUs for facial expression and 12 locations that account for head orientation and gaze. Two AUs from the FACS system are presented in Figure 3.4 (based on Ekman and Friesen, 1978).

AU 53: Head Up   **AU 54: Head Down**

Figure 3.4: The Action Units (AUs) that comprise a head nod movement.

By isolating the existence of movement in the AUs, specific forms of NVB and/or NVC, such as smiles, frowns, and so on can be determined. This is achieved by means of using a statistical algorithm, a Hidden Markov Model

(HMM) classifier, which automatically analyses the transformation from one AU to another in a sequence of video frames in order to model particular sequences of movements of, and around, given AUs. HMM classifiers are commonly used for modelling time series data, for examples of related studies see Avilés-Arriaga and Sucar (2002: 244).

As seen in Figure 3.4, when using the FACS system it is suggested that a consecutive combination of AU53, head-up, and AU54, head-down in any order from one frame to the next, has the potential to be classified as a head nod, following the HMM analysis. Consequently if, for example, AU54 is preceded and followed by the cessation of movement, or indeed any other AU, a 'no-nod' sequence is likely to be registered instead.

Another commonly used movement-based coding system is McNeill's gesture phase coding scheme, an illustration of which is depicted in Figure 3.5. This scheme allows the modelling of a range of bodily movements, beyond the head and face, predominantly concentrating on defining sequences of hand movement.

The only real drawback of such a movement-based scheme, as with the other schemes detailed above, is that they are intra-semiotic by nature (see the reference to Baldry and Thibault above). These schemes are designed to tackle movements alone. They are not fully integrated with a mark-up system tackling features of the spoken language, or indeed for marking up more semiotic aspects of gesture, relating the visual sign to a derived *meaning*. So although they are 'very precise in one or two modalities….. they generally do not cover the entire multimodal domain not the very fine-grained level of annotation required in every modality' (Blache et al., 2008: 110). Nonetheless,

it is important to note that such schemes can be integrated with others as part of a wider system of annotation, during a second parse of coding. Within a given research project or paradigm, specific schemes are often utilised to mark-up specific features and then combined within a wider framework for analysis.
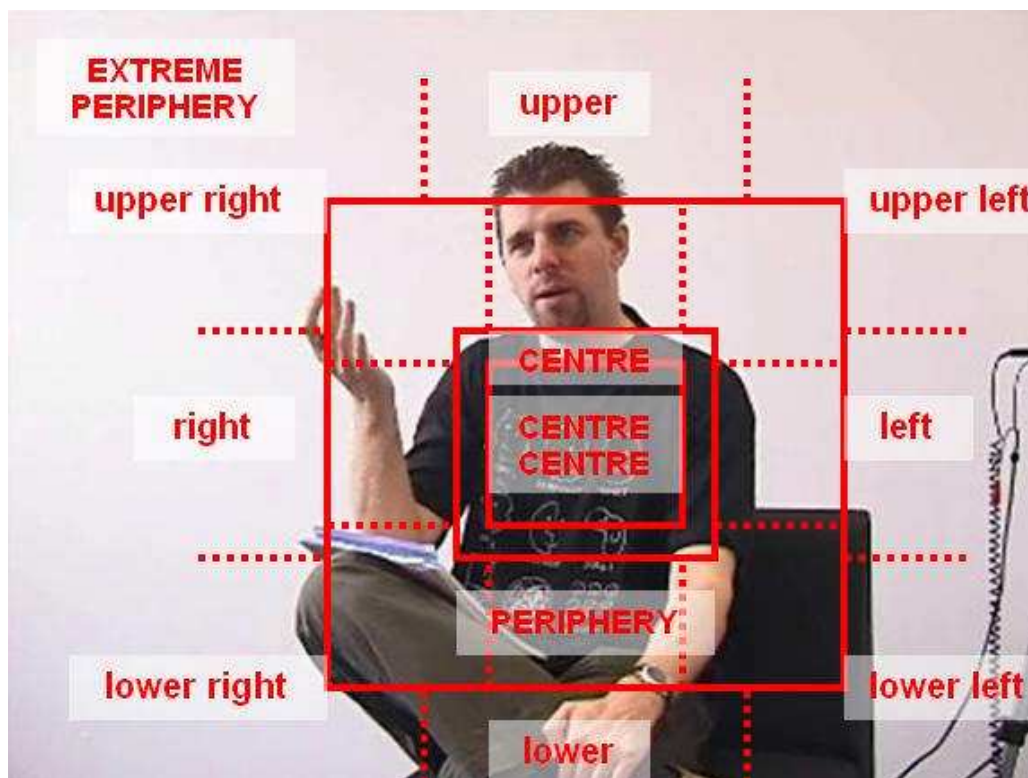
Figure 3.5: Division of the gesture space for transcription purposes, based on McNeill (1992: 378).

Other coding schemes which theoretically *are* equipped for dealing with both gesture and speech (a variety of schemes are discussed at length by Church and Goldin-Meadow, 1986 and Bavelas, 1994) merely tend to address specific typographic aspects of language and NVC. Examples include schemes designed to model sign language and/or facial expressions,

such as the HamNoSys[30] (Hamburg Notation System, see Prillwitz et al., 1989); the MPI Movement Phase Coding Scheme[31] (see Kita et al., 1997 and Knundsen et al., 2002) and DAMSL (Dialog Act Markup in Several Layers, see Allen and Core, 1997) which are designed to code gestures and signs which co-occur with talk. Similar to many of the systems for mono-modal corpora mark-up, these tend to function in XML.

Other schemes exist which allow coders to represent some elements of the basic semiotic and/or pragmatic relationship between verbalisations and gesture, i.e. focusing more on the meaning relationships between gestures and other concurrent interactive signifiers (early coding schemes of this nature are provided by Efron 1941 and Ekman and Friesen 1968; 1969). These annotate, for example, the occasions where gestures co-occur or not, with speech, and whether the basic discursive function of the gestures and speech which 'overlap', or are 'disjunct', or whether concurrent verbalisations and/or gestures are more 'specific' than the other sign at a given moment (for further details see Evans et al., 2001: 316).

Examples of coding schemes of this nature include one devised by Cerrato (2004: 26, also see Holler and Beattie's 'binary coding scheme for iconic gestures', 2002, and Allwood et al's MUMIN coding scheme 2007a, featured in Section 2.3.1.1 of Chapter 2). Cerrato's scheme was used to mark up a range HH and HCI conversations according to processes of feedback, distinguishing situations where feedback is 'given' (marked with Giv) from those situations where feedback is 'elicited' (marked with Eli) by means of

---

[30] HamsNoSys is a coding scheme for sign languages. For more information consult:
www.sign-lang.uni-hamburg.de/projekte/hamnosys/hamnosyserklaerungen/englisch/contents.html
[31] For more information on these tools and please consult the Max Planck Institute website
(MPI) at http://www.mpg.de/english/portal/index.html

both spoken and non-verbal contributions (so across the modalities), and not restricted to speech and/or gesticulation.

Regardless of the scheme used, it is important to note that little agreement exists across these different schemes. This is also true for current approaches to mono-modal coding and annotation (and transcription). So there are no conventionalised prescriptions that determine which behaviours to mark-up, how these elements are defined, and how they are physically annotated and integrated in the digital records of behaviour (in this case, the corpus database). Furthermore, there is little agreement on how these methods can best be integrated in order to cater for both spoken and non-verbal behaviours, that is, for the MM elements of discourse.

A priority in MM research is to draw up steps for generalised standards for this. Relevant schemes for the codification of visual and/or spoken data have recently been compiled by various researchers and research teams. Most notably, the ISLE project mentioned above, has started to lay the foundations for creating 'International Standards for Language Engineering' (Dybkjær and Ole Bernsen, 2004: 5), a 'coding scheme of a general purpose' to deal with the 'cross-level and cross modality coding' of naturally occurring language data (Dybkjær and Ole Bernsen, 2004: 5-8, also refer to Wittenburg et al., 2000). These standards, known as Natural Interaction and Multi-Modal Annotation Schemes (NIMMs), are designed to annotate 'spoken utterances, gaze, facial expressions, gesture, body posture, use of referential objects and artefacts during communication, interpersonal (physical) distance etc, and combinations of any of these' (Dybkjær and Ole Bernsen, 2004: 5). This is

with the aim to integrate these aspects to develop re-usable and international standards investigating language and gesture-in-use.

However at present, as with similar approaches, the NIMMs have not been formally presented to the research community and, furthermore, information concerning them is difficult to access, limiting the potential usability of this standardised scheme. Additionally, the ISLE standards have not been constructed specifically for linguists and this may have an effect on its adaptability for use in CL methodology. Despite this, the premise behind ISLE is a very real methodological requirement for 4th generation corpora. It is one which promptly needs to be addressed by corpus developers, as such global conventions are integral to the construction of high quality re-usable MM corpus datasets in the future.

As a final note, it should be emphasised that irrespective of the specific coding schemes and approaches used by a researcher or corpus builder, the fundamental importance is that they are both proficient and fully functional. Discussing the coding of qualitative datasets specifically, Edwards (1993: 21-23) suggests that coders, therefore, need to ensure that specific codes and schemes are 'systematically discriminable' (whether it fits a category or not); 'exhaustive' (ensuring all possible forms of a specific phenomenon are accountable) and 'systematically contrastive' (so that categories are mutually exclusive as far as possible). Ide offers similar suggestions, emphasising the need for consistency across the data streams and the need for the maximum processability of schemes for digital use (1998: 1-2).

3.5.3. Digital coding tools

Similarly with transcription, there are a plethora of software toolkits which support the digital encoding of MM datasets. Many of these integrate transcription functionalities with basic coding capabilities (including some of those mentioned in Section 3.4.2.2). The software tools Constellations[32] and Dynapad[33] are designed to specifically link and present pre-coded and pre-transcribed data, to allow subsequent analyses of the data to be undertaken. Other tools such as CLAN[34] and I-Observe[35] (see Badre et al., 1995: 101-113 for details), provide interfaces for coding and/or time-stamping video and textual data.

Finally, the Diver Project[36] (Pea et al, 2004); the Observer[37]; NVivo[38]; Atlas.ti[39]; ELAN[40] (see Brugman and Russel, 2004, used in the IFADV corpus); Mediatagger[41] from the MPI (codes are assigned using this tool, then

---

[32] Constellations is an 'event based' analysis tool which allows users to synchronise and time align multiple modes of data. More information can be found at: http://orion.njit.edu/merlin/tools/c25/index.html

[33] Dynapad is a multimodal visualisation (representation) tool, see http://hci.ucsd.edu/lab/dynapad.htm for further details.

[34] CLAN is a tool that allows for the coding and analysis of text, compatible with the CHILDES corpus and transcription database http://childes.psy.cmu.edu/clan/

[35] I-Observe is an ethnographic data collection and organisation tool, designed specifically for creating surveys, conducting polls and so on. For more information consult: www.apple.com

[36] Diver allows users to synchronise, view and play multiple video streams within a single resource. Integrated videos are called dives and may be explored by browser software. Diver is no longer available online, as it is currently being integrated with Dynapad (see footnote 33).

[37] The Observer is a commercially available tool, designed for coding and analysing observational data sets. The Observer can be purchased online from: www.noldus.com

[38] NVivo is a commercially available product that supports the alignment and analysis of multiple multi-media data streams. NVivo can be purchased online from: www.qsrinternational.com/products_nvivo.aspx

[39] Atlas.ti is a qualitative-based multi-media analysis tool. It is commercially available from: www.atlasti.com/

[40] ELAN is a multimedia analysis and representation tool which is available for free online, see: www.let.kun.nl/sign-lang/echo/index.html?http&&&www.let.kun.nl/sign-lang/echo/data.html

[41] MediaTagger is Mac based software which facilitates the codification of video data at different 'tiers'. For more information see: http://www.ldc.upenn.edu/annotation/database/abstracts/brugman.txt

are inputted into the software EUDICO[42], see Knudsen et al., 2002), and the NITE XML Toolkit[43] (see Carletta et al., 2003, used in the SAMMIE corpus) to some extent support the processes of the coding and annotation of text and/or videos, and also provide some facilities for data visualisation (see Section 3.6 for the application and presentation of data), thus integrating these facilities.

Again, the specific tool(s) chosen for use in a particular research project or study is very much reliant of the specific requirements of the end-user, and are thus chosen in light of their ability to fulfil the needs of the analyst. In relation to this, it is important to note that although the majority of the tools mentioned in this chapter are integrated with appropriate applications which allow for the transcription, coding, presentation and/or interrogation of MM datasets, they have not been designed specifically to help construct or host MM linguistic corpora. Therefore, they are somewhat limited in their usefulness for corpus-based interrogation and analysis of datasets.

However, currently there are two tools available which are specifically designed to support the annotation and analysis of multimodal linguistic corpora; ANVIL[44] (Kipp, 2001 and Kipp et al., 2007) and DRS (Greenhalgh et al., 2007, previously known as the ReplayTool, see French et al., 2006). Indeed, ANVIL was used when developing the CID and Fruit Carts corpora (see Figure 2.1 in Chapter 2), while DRS was used to develop the NMMC, as part of the DReSS project. Consequently, these tools are possibly the most

---

[42] EUDICO stands for the European Distributed Corpora Project based that the Max Planck Institute website, see: http://www.mpi.nl/world/tg/lapp/eudico/eudico.html

[43] NITE XML is a workbench of tools that allows for the annotation natural interactive and multimodal data. NITE XML can be downloaded for free from: www.ltg.ed.ac.uk/NITE/

[44] ANVIL is a frame accurate multimodal annotation and visualisation tool, available for free from: http://www.dfki.de/~kipp/anvil/

relevant to the aims of MM CL constructors since they have been built with this purpose in mind.

Both ANVIL and DRS allow users to construct time-stamped transcripts, align these with video and/or audio records, and to encode features of interest within and across each stream of data, within individual coding tracks. These coding tracks are thus tied, by time, to the video and transcript. In DRS, data records aligned with transcripts and coding tracks are visualised in a bespoke 'track viewer' (in the same way as the 'annotation' track in ANVIL), as depicted in Figure 3.6.
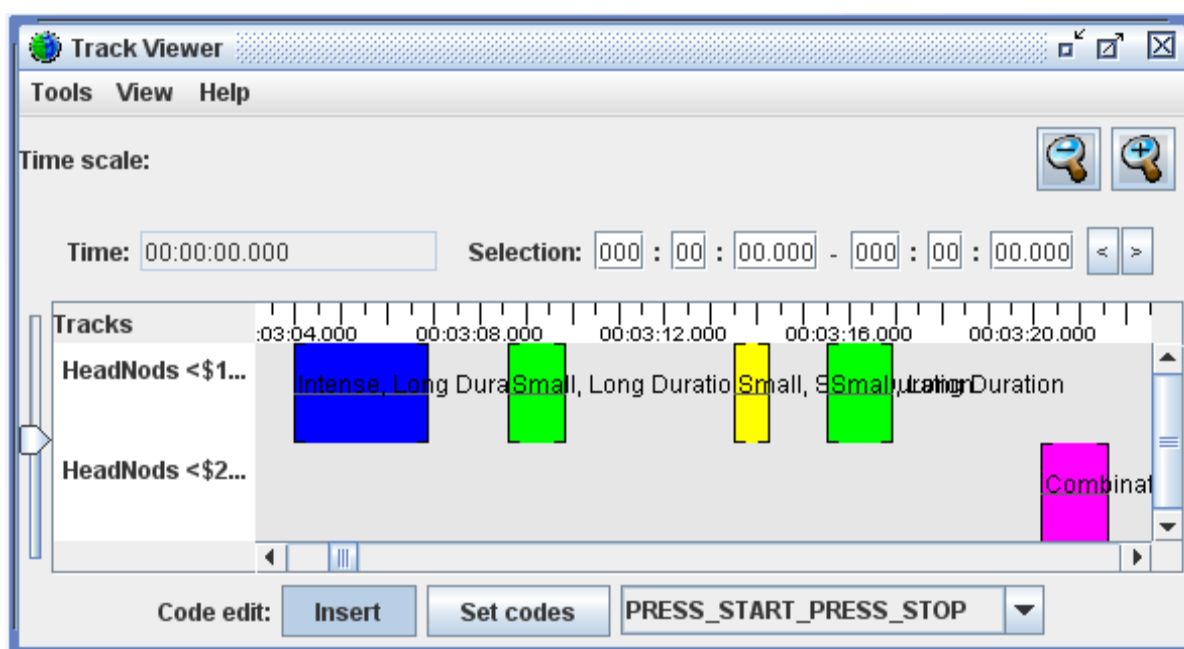
Figure 3.6: The coding 'track viewer' within the DRS environment.

This track viewer allows the user to add multiple track's comprising any form of time series data, providing the user with an accurate method for navigating and interrogating, potentially, large-scale datasets; across a

number of different speakers, and encoding a variety of different textual, gestural or extra-linguistic elements, as desired.

For example, Figure 3.6 represents an 'intense nod of a long duration' (i.e. type **D**; see Chapter 4 for further details) in **blue** in the track viewer, while a 'small nod of a long duration' is highlighted in **green**, and so on. The different colours used to denote these conditions and the adjustable size of the associated colour blocks provide an easy-to-use reference point for examining, in this case, the location of the head nod, the type used and the approximate duration of each. Coding can be undertaken using a right click utility within the track viewer to define start points of action and dragging or clicking the mouse to stipulate the end points.

3.6. Applying and presenting corpora

3.6.1. Key requirements

The final stage in corpus construction concerns the application and presentation of data. In other words, it seeks to address how corpora are presented to the end user, once data has been collected, transcribed and coded. The notion of the (re)presentation of data is heavily reliant on the software used by the corpus developers, as this determines how the data, including the raw video and/or audio files; transcripts and separate coding tracks; metadata; header information and so on, is arranged within the software's infrastructure. The software also determines how the data is navigated, searched and interrogated in screen. Again, as with previous stages of development, it would be preferable if the conventions used at this

stage were universal, however at present this is not the case as a range of different forms of corpus software exist.

Having said this, most current corpora are integrated with a key functionality which operates in a similar way across each individual database; a text concordancing tool. An example of a typical concordance output is seen in Figure 3.7 (taken from CANCODE).
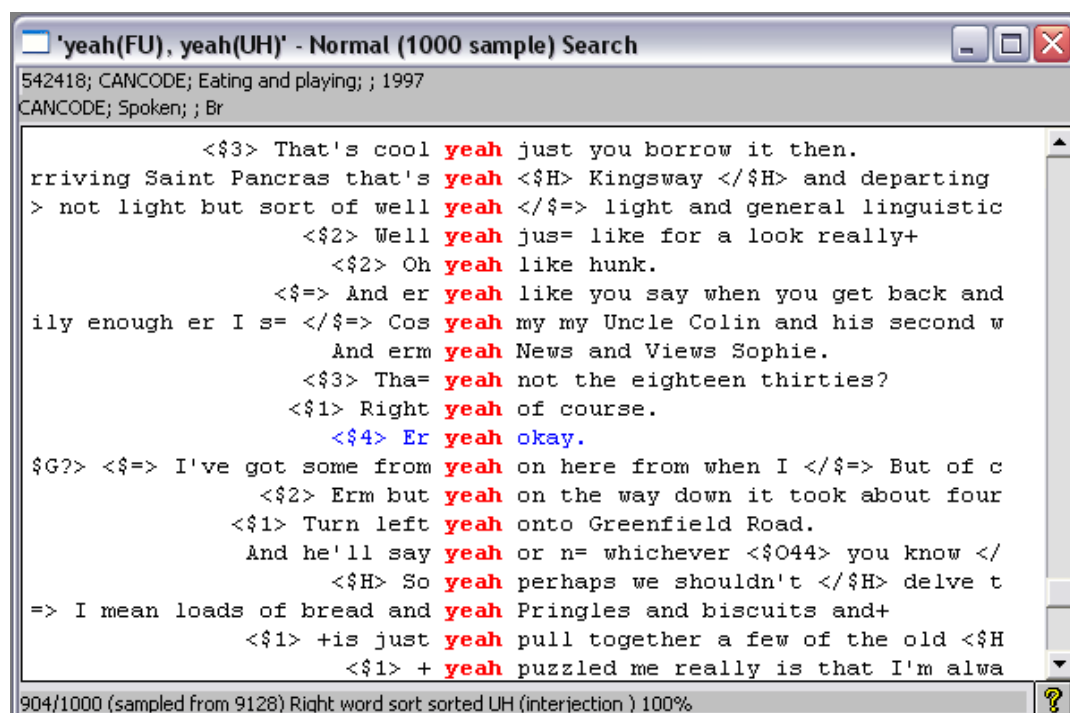
Figure 3.7: An example of 3rd generation corpus concordance outputs.

It is this which when coupled with search and word count facilities, allows the user to research statistical or probabilistic characteristics of corpora, together with exploring specific lexemes, phrases and patterns of language usage in more detail. At the click of a button, appropriate citations of speaker information, socio-cultural context of use and further details of the specific conversation in which each search term, line and/or turn occurs (as presented

in concordance output), can usually be accessed. Some of this is the information that forms part of the metadata content of the corpus.

The key limitation with such concordancers, however, is that they are only able to interrogate transcripts and text files, and not MM and/or ubiquitous datasets, as there is a scarcity of concordancers that deal with MM corpora specifically. For the advancement of 4th generation corpora it is vital that this void is filled and capabilities for conducting corpus-based searches of MM data are enhanced. However, this process is no mean feat as with the onset of MM, multi-media datasets present a whole host of technological challenges for the synchronisation and representation of multiple streams of information.

In an attempt to construct some guidelines for software which allow for the presentation and interrogation of MM datasets, in addition to the coding, organisation and management of such, the following key requirements were established at the start of the DReSS project. Although these principles act as benchmarks that were specifically constructed with the NMMC in mind, they can be seen to be valid beyond the remit of this corpus, and act as useful prescriptions for other MM corpora (see Knight et al., 2005: 12):

- *Multi-modal*: Allowing for the analysis and exploration of data from a variety of multimedia (sound and visual data) simultaneously, both within a single frame and a combined frame of reference when desired.
- *Accessible*: It should be integrated with a user-friendly interface to access and search specific frames or sequences of frames.

- *Proficient*: To be able to synthesise, tag, code and transcribe large quantities of MM datasets.

- *Flexible*: Allowing the interrogation of specific frames or sequences of data, as well as allowing the exploration of specific modes of data.

- *Systematic*: It should enable accurate and systematic searches and statistical analyses of spoken and visual records to be undertaken with ease.

3.6.2. Presenting multi-modal corpora in DRS

In light of these requirements, it should be noted that what sets DRS apart from ANVIL (and the other tools mentioned above) is that it is integrated with a fully MM search and concordancing facility for text and video data. Furthermore, it is also integrated with a facility that allows users to conduct basic text-based word frequency searches of corpora. So, in addition to providing the standard mono-modal concordance facilities seen in current corpora (as depicted in Figure 3.7), this MM concordancer allows users to search for gestural codes within the output. This provides an easy point of access for analyses of patterns of behaviour across the different modes, highlighting the tool's accessibility. No other multimodal analysis or annotation tool is equipped with this facility at present. For this reason, DRS currently exists as the most suitable tool for MM linguistic corpus development and presentation. An example of the concordancer search facility is seen in Figure 3.8.
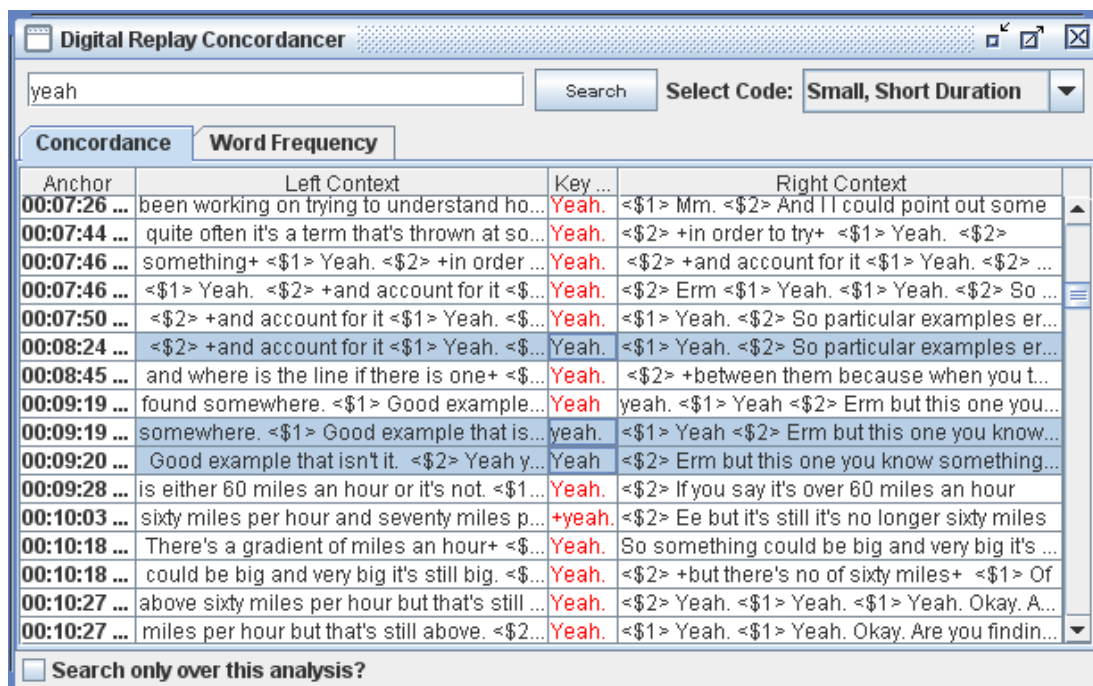
Figure 3.8: Exploring backchannel behaviour using the DRS concordancer.

In this figure, the standard text based concordances of *yeah* are presented. As the 'select code' box, in the top right corner, is enabled and a given gesture code is selected (in this case small nods of a short duration), relevant concordance lines are highlighted indicating where the search term and the specific coded gesture co-occur in close vicinity. In this case the figure indicates that where *yeah* is uttered and a small nod of a short duration is also enacted at some point within this turn.

Using the DRS concordancer, it is possible to search around the immediate environment of textual concordances using the right-click mouse facility. This enables the user to directly access the time-stamped segment of a transcript, and associated position in the text and video where specific events occur or where a particular lexeme is uttered.

At present, the tool does not allow for searches of specific gesture codes directly within the concordancer. Future releases will hopefully enable this line

of enquiry, for example, allowing users to search for specific gesture codes such as <NOD> and to calculate the frequency of these in the given text(s).

Since the current thesis is corpus-based and thus relies on these sorts of concordancing applications for research, DRS is invaluable. However, given that the concordance tool and related functionalities including the word count facility, are still relatively new components, its reliability may be questioned because extensive testing of this functionality has yet to be carried out. As a result, both the case study and extended five-hour datasets used in this thesis do not utilise this frequency tool, although it noted that this application is sure to be invaluable for MM corpus research. Since no alternative MM concordancers or frequency tools exist at present, the majority of the searching and counting conducted here has been undertaken manually (see Chapters 4 and 5 for further details).

3.7. Summary

This chapter has explored the processes of developing MM corpora, drawing on a wide range of issues and methodological considerations that need to be addressed; from the process of recording MM conversational data through to its' representation and re-use. Although this developmental methodology is by no means definitive, it has provided a context to MM CL research, by outlining some of the key practical, technological and ethical questions that are faced.

Effectively, this research provides a background for the second focus of this thesis; the actual implementation of such corpora. Chapter 4 examines this matter in more detail. The chapter outlines a refined approach, a framework, one which enables accurate and relevant analyses of MM corpus

datasets to be undertaken, taking ten minutes of case study data as a means for doing this. This analytical framework is, in turn, used as the basis for analysing 5 hours of NMMC data in Chapters 5 and 6.

Chapter 4: Multimodality and Active Listenership

4.1. Introduction

This chapter aims to do the following:

- Outline a methodological approach for the analysis of spoken and non-verbal behaviour (i.e. signals of active listenership) in MM corpora. To achieve this, the chapter investigates the best methods to implement in order to undertake the following processes:
    - How to *define*, *extract* and *encode* spoken and non-verbal behaviours.
    - How *patterns* in behaviours within and across the data streams are determined and qualified.
- Illustrate this revised approach in operation by using the case study data, in order to determine the adequacy and practicality of the approach. This line of enquiry will trial how well the approach works in practice, highlighting problems faced throughout.

The chapter highlights the requirements for constructing a revised CL based approach to the analysis of MM corpus data. This is to provide a framework that not only caters for spoken behaviour in conversation, but is also viable for use when analysing features of gesture-in-talk. The chapter concentrates on outlining a principled 'manual' approach to the analysis, presenting each stage from the transcription and presentation of the raw data, through to the extraction and definition, quantification and analysis of features of interest. In

essence, it concentrates on defining and demonstrating what a user can actually *do* with MM corpora data once it has been collected. This is achieved by means of conducting a case study analysis of patterns of spoken and non-verbal backchanneling behaviour in a short extract of a single supervision session from the NMMC. This case study functions as the pilot study for the thesis.

4.2. Approach

4.2.1. Data used

The following study performs an in-depth analysis of a ten-minute excerpt of a face-to-face, human-to-human, dyadic conversation. The excerpt has been extracted from a forty-five-minute video of an MA supervision session involving a male supervisor (<$M> hereafter) and female supervisee (<$F> hereafter), both of whom are British nationals. This is the first dissertation supervision between the participants who, prior to this meeting, had only conversed on a few brief occasions. The supervision was randomly selected from the NMMC, and the excerpt was extracted from the middle of the session, between minutes 15 and 25.

The reason for only selecting 10 minutes of data for this case study was that it is designed to test whether the methods and approaches set forth in this chapter, and in those preceding it, are appropriate for the systematic exploration of MM data. Before attempting to analyse larger data sets, it is logical to make sure that the optimal ways of tackling the data at every stage of the analysis are established.

Biber contends that a mere 1000 words of corpus data are often sufficient for conducting 'basic' linguistic analyses (1990, 1993). Given that this excerpt comprises 2200 words of data, as a 'basic' pilot study analysis, this sample is indeed of an adequate size to achieve this. In further support of this claim, Flowerdew suggests that provided 'that there is a sufficient number of occurrences of a linguistic structure or pattern' (based on Flowerdew, 2004: 25) to allow for the aims and objectives to be explored, the data sample is 'sufficient'. As both spoken and non-verbal backchannels are characteristically relatively frequent in everyday conversation (refer to comments made in Chapter 2), it can hypothesised that a ten-minute sample of dyadic conversation data should actually provide more than enough stimuli for conducting fundamental investigations into these phenomena.

4.2.2. Current corpus analysis conventions

In contemporary corpus-based research, the analyst is typically concerned with exploring 'the patterns of language, determining what is typical and unusual in given circumstances' (Conrad, 2002: 77). These patterns are identified following a 'quantitative assessment' of given phenomena, one that can be linked to early linguistic research and, in particular, the work of Zipf (see Chapter 3 for further details).

The analysis of modern computerised corpora is generally structured according to one or more of the following research models (Leech, 1991: 20):

- *Data retrieval model*: Machine provides the data in convenient forms. Human analyses the data.

- *Symbolic model*: Machine presents the data in (partially) analysed form. Human iteratively improves the analytic system.

- *Self organising model*: Machine analyses the data and iteratively improves the analytic system. Human provides the parameters of the analytic system and the software.

- *Discovery procedure model*: Machine analyses the data using its own categories of analysis, derived by clustering techniques based on data. Human provides the software.

Of these four approaches, the data retrieval model relies on the most human interaction for the analysis of language data. Conversely, the symbolic and self-organising models use an increasing amount of computing power, and rely less on the work of the human analyst. The fourth model listed here, the discovery procedure model, uses the minimal amount of human interaction within the analysis process. Instead, it relies solely on sophisticated computerised software.

The more manual models of analysis, i.e. those at the top of the list, are often best suited to 'hypothesis-driven' research where 'a specific linguistic research question, which is identified at an early stage in a research project, leads to the collection or selection of a corpus and some phenomenon is investigated using that corpus' (Rayson, 2003: 1). In such a research paradigm, the analyst has a lot of control over the examination of the data and relies less on the analytical power of modern corpus software. In contrast, the

latter approaches (the self organising model/ discovery procedure model) are best conceptualised as being more 'data-driven' insofar as the analyst is 'informed by the corpus data itself and allows it to lead us in all sorts of directions' (Rayson, 2003: 1).

It is difficult to highlight exactly which approach is most suited to the aims of this thesis. As seen in the previous chapter, the *developing* nature of MM CL utilities means that the interrogation of MM data relies heavily on a pre-analysis stage. This involves the actual *definition* and *labelling* of gestural phenomena, since this process is novel to conventional CL-based research. Thus, for this thesis, this pre-analysis involves demarcating where backchanneling head nods exist and assigning preliminary codes to these, based on characteristics of the movement shape of the nod and/or the functions they are seen to fulfil in the discourse (see section 4.2.3.2 for further details). Therefore, this part of the process is heavily data-driven, and closer to the data-retrieval model than the other end of the research spectrum.

It is only once these codes and/or categories have been established and every instance of head nod behaviour has been marked accordingly, that the analysis proper can commence. During this phase of analysis, a more hypothesis-driven approach (closer to a discovery procedure model) is taken in order to answer more specific questions regarding the nature and use of these behaviours, and their interaction with spoken forms (as outlined in the introduction).

Consequently, it can be effectively argued that a mixed-method approach is required in this thesis. This combines some characteristics of the data-

driven and hypothesis-driven approaches, and similarly techniques of the data retrieval model and the discovery procedure model.

4.2.3. A new methodological approach for analysis

4.2.3.1. Detecting and defining backchannels

Before approaching the analysis proper the issue of the pre-analysis requires further discussion. The following stages require to be considered and/or undertaken as part of this phase for the purpose of MM data analysis (based on Gu, 2006):

1. Multi-modal text has to be digitised and becomes processable by the computer.
2. Non-discrete streams of flowing images have to be segmented into discrete units that correspond to the analytic units of the content.
3. A metalanguage has to be constructed to annotate the segmented units.

The first stage, digitisation, concentrates on transforming real-life linguistic performance into 'data'. This is a fairly straightforward process for MM corpora, especially given the capabilities of technology in the modern digital age (and for other reasons discussed in Chapter 3).

In terms of digitising the data for use in this thesis, mini DV cameras with external stand-alone microphones were used to record and automatically store audio and visual records of conversational episodes (refer back to

Figure 3.2 in Chapter 3 for details of the specific set-up plan used). These were stored in a compressed MOV format; the format favoured by the DRS tool. Additionally, the textual records, i.e. the transcriptions of these conversational episodes, were also digitised for use. They were time-stamped, turn-by-turn, using the Transana software to temporally align the textual script with the video stream. This process allowed the transcript to be easily navigated.

In theory, the process of segmentation, the second step identified by Gu (2006) is relatively easy to address with textual transcripts of data. This is because the words and phrases defined by typical language conventions are easily defined, orthographically, in terms of discrete units. So, for this case study and the main study (refer to Chapter 6), the 'analytic units' of these spoken records are, specifically, lexical backchannels.

However, the physical segmentation of these spoken analytic units is a more challenging process. Although concordancing software, such as Wordsmith (see Scott, 1999[45]) and DRS, can easily determine the raw frequency counts for given word forms, they are not usually adept for searching semantic or pragmatic categories of linguistic phenomena, such as backchannels. This is because a corpus may be grammatically or syntactically tagged (see Leech, 1991), but few corpora are semantically tagged for instances of backchanneling behaviour. Therefore, when searching for specific forms of backchanneling behaviour, rather than simply typing 'backchannels' into the search box, the analyst is required to manually search for each possible form of backchanneling behaviour in turn.

---

[45] Wordsmith Tools is a lexical analysis toolkit developed by Mike Scott, published by the Oxford University Press. For more information and to purchase Wordsmith Tools version 5.0 see: http://www.lexically.net/wordsmith/

Given this, the following steps were taken in the case study as a pre-analysis phase for the segmentation of spoken backchannels. These are semi-automatic as although digitally based, they do require a certain level of laborious manual processing:

A. Searching for the most common 'simple' and 'series' backchannel forms (Oreström, 1983 and Tottie, 1991, specific examples of these forms are provided in Chapter 2) in the corpus, using the DRS concordancer (further details in 2.2.3).

B. Searching for possible derivations of the forms identified in A, and searching for other less common forms of spoken backchannels, using specific forms noted in past research (as documented in Chapter 2).

Problems concerning the consistency of transcription conventions, as faced during the transcription phase of corpus development (see Chapter 3) can affect the accuracy with which these searches are undertaken, as a range of different spellings for certain backchannels forms exist. So, for example, an instance of the minimal form *mmm* may be transcribed as *m*, *mm* or *mmmm* by a range of different analysts. Arguably, none of these forms are more 'accurate' than the other, since such an utterance is effectively non-standard, thus obviously no fixed spellings exist for this. Consequently, to run accurate frequency counts and corpus searches, it is necessary that derivational spellings of the specific verbalisations are accounted for if there is a likelihood that inconsistencies of such exist in the corpus. Otherwise many instances of backchanneling behaviour will be overlooked and results generated by

subsequent analysis are likely to misrepresent the data. Since the NMMC was built by a small team of researchers, all of whom were aware of the importance of consistency in the spelling of these units, this matter was addressed and relevant guidelines drawn up when the data was initially transcribed.

As identified in Chapter 3, the challenges faced when attempting to segment non-verbal behaviours is far more complex and multifaceted than for spoken counterparts. While the crux of a MM CL approach necessarily requires that gestures are converted into 'discrete units that correspond to the analytic units of the content' (Gu, 2006: 146, see section 4.2.3), they are by nature, paradoxically, 'non-discrete' (Gu, 2006: 146) and thus difficult to convert into units (refer back to Chapters 2 and 3 for further details). So, through the exploration of gesture-in-use, an attempt is made to essentially define and model behaviours that contradictorily are not lent to being defined. However, it should be noted that although this process of segmentation and definition is problematic, it is unavoidable if these behaviours are required to be explored in more detail. More appropriate and/or accurate, alternative, strategies for this do not currently exist.

In an attempt to segment non-verbal units in this case study, the following stages were negotiated:

i. Defining instances of head movement.

ii. Determining whether a given head movement (from i.) is a head nod.

iii. Determining whether a head nod (as determined in ii.) is a backchanneling head nod or not.

These three steps foreground a manual approach to head nod definition. The first two steps involve marking up the actual existence of head movements in discourse. This is undertaken by means of watching and re-watching the video recording and utilising information about the kinesic properties (movement structure) of head movements in order to assist the definition of 'nod' and 'no-nod' sequences. These comprise, as crudely defined, the up-and-down sequence of the head motion, whether it does or does not occur; in any order. This manual method is necessarily interpretive and inferential, as is solely dependent on the subjective opinions of the analyst. Alternative, (semi)automated approaches for this process are discussed extensively in Chapter 5.

The third stage of this segmentation process is arguably the most challenging. There is currently no regimented way to define whether a given nod, with a particular movement structure or intensity positioned at a given location in talk, is likely to be a backchannel any more or less than others. This is because, as yet, little is known about the ways in which individual backchanneling nods, nods of a particular form, function in talk.

This phase is again inferential and, as defined in Chapter 2, is best facilitated by, the exploration of the discursive co-text and context of the head-nods, for example. So, theoretically speaking, if a nod follows a presupposition or polar-type question, the nod is to be classified as a specific response to that question rather than a backchannel. Whereas, if the nod is administered mid-way through talk or at 'completion points' (refer to section 2.4.2.3 in Chapter 2, Blache et al., 2008: 114) and/or TRPs (as with spoken

forms), it is possibly more appropriate to define it as an example of non-verbal backchanneling instead.

Once the head nods have been defined, it is beneficial to categorise them according to their basic movement features and their physical location in context of, and in relation to, spoken forms of backchanneling behaviour, since this relationship is of a primary concern to the thesis. Firstly these backchanneling nods can be categorised according to their specific movement sequences. To establish a relevant system for these classifications it is logical to start with five simple *types* of nods, as follows:

Type A: Small (nonchalant) nods with a short duration.

Type B: Small (nonchalant), multiple nods with a longer duration than type **A**.

Type C: Intense nods with a short duration.

Type D: Intense and multiple nods with a longer duration than type **C** nods.

Type E: Multiple nods, comprising of a combination of types **A** and **C**, with a longer duration than types **A** and **C** nods.

Files 4.1 to 4.5 on the data disk provide video examples of these nod types, using data taken from the NMMC data (for nod types **A** to **E** inclusive).

For the purpose of this system, movements are categorised from 'nonchalant' and 'short' through to 'intense' and 'long', and a combination of these. The intensity of nods is defined in terms of the amplitude of the head movement, the physical size of the movement in the head-up or head-down

motion. Therefore, nods which appear to exact a more physically extreme motion are likely to be classified as types **C** or **D** (or **E**). Whereas nods with slight movement in the up or down direction are classified as being more nonchalant, so types **A** or **B** (or **E**).

In addition to the physical shape of these nods, the 'location' of these forms of backchanneling nods can then be broadly sub-categorised, according to the following two groups: those nods which co-occur with a spoken backchannel and those that are used without spoken backchannels.

Again, the reliability and accuracy of these classifications is solely reliant on the skills of the analyst, who is required to inspect and closely re-inspect each nod and determine the most appropriate category for that item. Indeed in many cases, it is possibly easier to determine what the item is not, before narrowing it down to an appropriate category. In large-scale studies, 'multiple passes' of manual assigned annotations of data should be undertaken as an act of quality control (Strassel and Cole, 2006: 3), with annotations and codes being checked and double checked by a range of different coders. This is to ensure consistency and inter-rater reliability, and can be undertaken by means of the following tests (Cerrato, 2004: 27):

1- Stability or invariance test which checks whether the same coder varies his/ her judgements over time.

2- Reproductability test or inter-coder variance which checks the agreement of two coders.

3- Accuracy test which compares the coding produced by these two coders to the standard, if the standard is available.

However, in terms of the current thesis, it was obviously not possible to do this, as there was only one person coding. Nevertheless, such measures of reliability should be integrated into the design and analysis of MM corpora.

4.2.3.2. Coding and marking-up phenomena

The final stage of Gu's process of analysis is to 'develop a meta-language and annotate the segmented units' (Gu, 2006), i.e. to construct a system for codifying and marking up specific backchanneling elements in the data. In terms of spoken backchannels, building on steps A and B above, essentially the concern is with accomplishing the following:

C. Determining what discourse functions these forms possess, based on where and when the token in used within the structure of the discourse, and according to the sense of meaning it creates.

The classification of the discourse functions possessed by the spoken backchannels is undertaken in accordance with O'Keeffe and Adolphs' (2008) functional cline, outlined in Chapter 2.

Chapter 2 considered that even when presented with a common and frequent backchannel such as *yeah,* it can be difficult to ascertain whether it is functioning as a CON or CNV at any given moment in time. In consequence, when assigning the functions it is vital that the coder (analyst) considers spoken forms of backchannel more widely, in the context of the remainder of the conversation.

In other words, it is necessary to take note of what comes before and after the backchannel in a sequence of talk; its co-text and contextual features (pauses, particular statements, questions and so on), whether it occurs at a TRP or not, and/or to monitor particular prosodic patterning of spoken forms (the manipulation and replaying of the audio stream in DRS facilitates this, alternatively particular 'problem' cases can be examined in depth using the Praat[46] phonetic software, Boersma and Weenink, 2005; outputs from Praat are also used in the VACE, IFADV and CID corpora). By examining each of these features, it is possible to get a better indication of whether a given form is: (a) functioning as a backchannel or not and (b) to map out the particular function the backchannel is adopting in discourse.

Although in Chapter 2 it was shown that similar functional classifications for non-verbal backchannels do exist (for an example see Maynard, 1987), at present there is no indication across the literature, of how these functions relate, if at all, to specific semiotic forms of backchanneling nods. So it is difficult to actively classify particular nods with ease, using such systems.

Given this, it is invaluable to explore the relationship between spoken and non-verbal backchannels, and to extrapolate patterns from the ways that each variety is used. This is in order to model some of the relationships between specific forms and functions across the visual and vocal stimuli, and to develop an understanding of the role and nature of non-verbal backchanneling behaviour in naturally occurring discourse.

---

[46] Praat is a freely available fine grained audio analysis tool. See:
http://www.fon.hum.uva.nl/praat/

4.2.3.3. Presenting data

Once the specific instances of backchanneling have been defined, it is necessary to physically mark them up for use in the analysis. For this purpose, the following characteristics of the data require annotation:

- Backchanneling nods: *Location* in the context of discourse and *Gesture shape,* based on kinesic properties
- Spoken backchannels: *Lexical form* and *Discursive function* (according to O'Keeffe and Adolphs, 2008)

As identified in Chapter 3, there are a variety of different ways to represent such information in transcript form. In other words a variety of different graphical and 'abstracted symbolic representations' (Saferstein, 2004: 213) can be used to mark the location and shape of head nods in a textual record. For example, Streeck (1994: 241) uses horizontal square brackets under an utterance, i.e. in a separate line in the transcript, 'to indicate the extension of a gesture' (Norris (2004: 112), whereas Saferstein (2004: 213) marks gestures in parenthesis.

Since the present study is dealing with a finite range of gesticulations, only a basic approach for annotating non-verbal and spoken backchannels in the transcriptions is required. Figure 4.1 shows the approach used, using an extract of the data examined in this chapter. Refer to file 4.6 on the data disk for the complete transcript of the case study data.

In Figure 4.1, instances where backchanneling head nods occur without concurrent spoken backchanneling forms (i.e. in 'isolation'), are underlined,

the colour used to do this depends on the identity of the 'nodder'. On the other hand, when spoken backchannels are uttered in isolation (i.e. without concurrent head nods), the text of the transcript is highlighted in a colour representing the functional classification of the given lexeme(s) or string.

TRANSCRIPT		KEY
<$1> Yeah. So you are looking at	c	**Verbal Backchannels (B-Cs):**
<$=> thousands and thousands of		Continuers (CON)
<\$=> well+	a	Convergence Tokens (CNV)
<$2> Yeah.		Engaged Response Tokens (ER)
<$1> +you're not up into the millions yet are you.		Information Receipt Tokens (IR)
<$2> I think I am going to have to	a	**Head Nods:**
narrow it down to a qualitative study	a	___ – Nod with speech: <$1>
<$1> Right. Cos otherwise you're gonna need some sort of concordance+	c	▨ – Nod without speech: <$1>
		___ – Nod with speech: <$2>
<$2> Yeah.		___ – Joint Nods
<$1> +or some sort of program		**a,b,c,d,e** = Head nod type
<$=> that <\$=> that+	b	**Speaker <$1> denoted in grey,**
<$2> and I think+		**<$2> in green**
<$1> +well identifies metaphor that		
<$H>don't exist really <\$H>		
<$2> <$H> not very much <\$H>.		
<$1> Erm yeah+		

Figure 4.1: Guide to mark-up and transcription conventions used in the case study.

Instances where spoken and non-verbal backchannels co-occur are represented as text that is both highlighted and underlined. Finally, text that has an absence of mark-up (i.e. with no segments underlined or highlighted) indicates when words are uttered without a spoken backchannel or backchanneling head nod being performed.

By underlining the approximate points in the discourse where head nods co-occur with speech or at positions of 'gesture alone' (Evans et al., 2001), and by highlighting the spoken backchannels, it is possible to present both

spoken and non-verbal forms together in the same transcript. This contributes to the greater manageability of the analysis of data within and across the different 'media'. Kress and van Leeuwen (2001) regard this to be a critical requirement for MM analysis.

4.3. Case study results

4.3.1. Specifying the areas of focus

Since this case study functions to pilot an approach for analysis, the results given here simply demonstrate the types of enquiry relevant to backchannel research, rather than define specific patterns of behaviour and/or assign meaning. However, Chapters 6 and 7 build on this enquiry. Therefore, the following analysis merely focuses on exploring elements of the location and gesture shape of nods as well as the lexical form (and structure) and discursive function of spoken backchannels; all in terms of frequency. The focus is placed on mapping the occurrence of backchanneling phenomena according to these characteristics, and defining relationships in the co-occurrence, rather than providing detailed statistical testing of this (see Section 6.2.3 in Chapter 6 for further discussions on this matter).

4.3.2. Spoken backchannel behaviour

4.3.2.1. Overview

Appendix 4.1 provides a breakdown of the different forms and associated functions of the spoken backchannels taken from the excerpt. The combined frequencies for these, and whether they co-occur with nods, are shown in Figure 4.2 (also see Appendices 4.2 and 4.3):

	Speaker		TOTAL
	<$M>	<$F>	
Spoken Backchannels with Nods:	24	23	47
Spoken Backchannels without Nods:	21	5	26
	45	**28**	**73**

Figure 4.2: A table showing the breakdown of frequency counts of spoken backchannels in the excerpt.

This table identifies that a total of 73 spoken backchannels are used in this excerpt, around 62% of which are spoken by <$M>, the supervisor (45 times, to the 28 instances by <$F>). This disparity can perhaps be justified by the fact that <$M> speaks more frequently than <$F> in the case study data. Of the 2156 words of the excerpt, 1401 were spoken by <$M>, whereas only 755 were spoken by <$F>. Interestingly, this equates to a constant rate of spoken backchannel use across both speakers, as the relative rate at which these speakers use spoken backchannels to words is *circa* 1:50 (45/2156 = 1:48 for <$M> and 28/1401 = 1:50 for <$F>). This suggests a natural constant in this behaviour, although the extent of this can not be fully determined in such a small data sample, and thus is explored in greater detail in Chapter 6.

Figure 4.2 also indicates that while both participants are more likely to use spoken backchannels with concurrent head nods than without, the proportion of usage for each of these states differs dramatically between the speakers. <$M> uses 53% of spoken forms with nods and only 47% without (i.e. 24 to 21), whereas <$F> uses a remarkable 82% with and only 18% without (23 to 5).

4.3.2.2. Lexical form

There are a large variety of different lexical forms of spoken backchannels used in this dataset, although many of these are used on only one occasion by one or other of the speakers (refer to Appendix 4.2 for specific details). Figure 4.3 charts the most frequent of these forms across the case study data excerpt; those with a frequency of >2 for each individual speaker:

Form	Frequency		Structural Type
	<$M>	<$F>	
Yeah	12	19	Simple
Right	10	0	Simple
Okay	3	2	Simple
Yeah yeah	4	0	Double
Erm	1	2	Simple
Oh yeah	2	0	Complex

Figure 4.3: The most common forms of spoken backchannels in the excerpt.

Figure 4.3 indicates that the most commonly used lexical form across both speakers is *yeah,* supporting the findings of past research which was discussed in Chapter 2. This is closely followed by *right* for <$M>, although this response is not used at all by <$F>. *Okay* and *erm* are also used by both speakers, but with relative infrequency.

In addition to form, this figure also classifies the 'structural type' of these commonly used backchannels. By examining the structure, variety in the forms of spoken backchannels that are used can be explored; depending on whether they are of a *simple*, *double* or *complex* form (refer back to Chapter 2, also see Oreström, 1983 and Tottie, 1991).

Overall, the figure suggests that simple, rather than double or complex forms of backchannels are the used most frequently in this data. In total, and this includes all instances of spoken backchanneling, not simply the most common forms, simple spoken backchannels are used 28 times by <$M> and 24 times by <$F>, whereas double and complex forms are used 4 and 12 times by <$M>, and 0 times and 5 times respectively by <$F>. However, there exists a wider range of different varieties of complex forms than simple forms in this data (i.e. of different lexical structures). <$M> uses 11 different complex forms, 1 double and 6 simple, whereas these figures are 5, 0 and 4 for <$F> respectively.

4.3.2.3. Function

Figure 4.4 shows the total frequency counts for each spoken backchannel functioning as CON, CNV, ER and IR token in the excerpt, occurring with and without concurrent backchanneling nods:

| | | Speaker | | TOTAL |
		<$M>	<$F>	
Function	CON	11	12	23
	CNV	11	13	24
	ER	7	2	9
	IR	16	1	17
		45	28	73

Figure 4.4: Frequency counts of spoken backchannel functions in the excerpt.

Figure 4.4 demonstrates that the most common discourse functions of these spoken backchannels are CON and CNV tokens. This is true of 23 and 24

spoken backchannels respectively, of the total of 73 instances seen. The table indicates that there is no real marked difference between the frequencies with which these functions are used across the two speakers. However, there is a more marked difference in the use of ER tokens, with <$M> using these tokens on 7 different occasions and <$F> on only 2 occasions. This amounts to 16% and 7% of the total number of spoken backchannels used by these respective speakers.

Use of IR tokens is shown to be even more inconsistent across the speakers. In total 17 spoken backchannels functioning as IR tokens are evident, 94% (16) of which were spoken by <$M>. This accounts for 41% of the total number of occasions on which this speaker uses spoken backchannels in this excerpt. Of these, 75% (12/16) were the response *right*, which is not used as a backchannel by <$F> at all. This suggests that a mere 6% of the total IR tokens used in the sample (1 from 17) were uttered by <$F>, amounting to around 4% of the total number of spoken backchannels used by this speaker.

4.3.2.4. Spoken backchannels with(out) concurrent nods

Figures 4.5 and 4.6 tabulate the frequency with which spoken backchannels and associated discourse functions occur with (Figure 4.5), and without (Figure 4.6) backchanneling head nods in the case study data (also see Appendix 4.3).

		Speaker		TOTAL
		<$M>	<$F>	
Function	CON	5	9	14
	CNV	6	12	18
	ER	4	1	5
	IR	10	0	10
		25	22	47

Figure 4.5: Concurrent spoken and non-verbal backchannels- a breakdown of discourse functions.

		Speaker		TOTAL
		<$M>	<$F>	
Function	CON	6	3	9
	CNV	5	1	6
	ER	3	1	4
	IR	6	1	7
		20	6	26

Figure 4.6: Spoken backchannels without concurrent head nods- exploring discourse functions.

The most striking difference between Figures 4.5 and 4.6 is that <$F> is shown to use both CON and CNV tokens with concurrent nods far more frequently than without nods (75% and 92% of occasions, respectively) whereas there are no dramatic differences in these frequencies for <$M>. Overall, <$F> is shown to use all forms of spoken backchannels without concurrent nods far less frequently than with nods (6 to 22 times), whereas for <$M> this rate is more stable (20 to 25 times).

4.3.2.5. A focus on 'yeah'

As an extension to these explorations on spoken forms, it is also interesting to look in more detail at the ways in which particular lexical forms are used as backchannels. Since Figure 4.3 identified that most lexical forms are relatively infrequent within such a small data sample, only *yeah* is focused on here. However, a wider range of forms are explored in Chapter 6.

There are a total of 31 uses of *yeah* in the case study data (<$M> = 12, <$F> = 19), 7 of which occur without concurrent nods (<$M> = 6, <$F> = 1), 22 with nods (<$M> = 4, <$F> = 16). In other words 50% (6 out of 12) of the *yeah's* spoken by <$M> occur with nods, whereas for <$F> this is 84% (16 out of 19); so in a mere 3 instances *yeah* is used without a nod for <$F>.

Figure 4.7 charts the frequency with which *yeah,* functioning as a CNV, co-occurs with and without nods in the excerpt:

Nod Type		Speaker		TOTAL
		<$M>	<$F>	
	A	0	3	3
	B	1	2	3
	C	1	3	4
	D	0	1	1
	E	2	1	3
		4	9	13

Figure 4.7: Frequency counts of *yeah* functioning as convergence tokens, and co-occurring types of backchanneling head nods (from the case study data).

On the other hand, Figure 4.8 maps the use of *yeah* as a CON token, with and without concurrent nods:

		Speaker		TOTAL
		<$M>	<$F>	
Nod Type	A	1	2	3
	B	1	1	2
	C	0	5	5
	D	0	0	0
	E	0	1	1
		2	9	11

Figure 4.8: Frequency counts of *yeah* functioning as continuers, and co-occurring types of backchanneling head nods (from the case study data).

These figures indicate that, for <$F>, when functioning as a CNV and a CON token, *yeah*, co-occurs most frequently with head nod types **A** and **C.** Although type **B** nods also frequently co-occur with the spoken backchannel *yeah* functioning as a CNV.

Figure 4.7 and appendix 4.2 reveal that *yeah* is used as a CNV with a nod on a total of 11 times (with 2 from <$M> and 9 from <$F>). In comparison, it is used as a CNV without a nod on only 4 occasions (3 from <$M>, 1 from <$F>). Similarly, Figure 4.8 highlights that *yeah* is used as a CON with a nod a total of 13 times (<$M> = 4, <$F> = 9), while it is used without a nod on only 3 (<$M> = 3, <$F> = 0) occasions. So <$F> uses *yeah* much more across the data, but more so with nods, whereas <$M> uses it more frequently without nods (refer to Appendix 4.2 for a breakdown of these results).

These results are interesting, but not altogether surprising. As although nod type **C** is described as being 'intense', whereas type **A** is 'nonchalant', both nods are described as being of a short duration, in the same way that *yeah* itself is phonemically short and monosyllabic.

4.3.3. Non-verbal backchannel behaviour

4.3.3.1. Results

It is important to note that while Figure 4.1 – 4.8 chart, where relevant, episodes in which a *single* spoken backchannel utterance co-occurred with a single nod, this is not always the case in discourse. In other words, a nod may instead be used across a number of turns where individual spoken forms are administered, and/or the same, single nod may be used with various different spoken forms of backchanneling behaviour. This provides justification for the reasons the case study has examined spoken forms with concurrent nods, and non-verbal forms with concurrent spoken backchannels separately, and explains why there is an apparent disparity in the number of nods documented from each perspective. Figures 4.9 – 4.12 chart backchanneling nods that co-occur with spoken forms across each of these states, in order to account for all of the ways in which nods are used with spoken forms.

Figure 4.9 identifies the frequency of backchanneling head nod use in the case study data, detailing then number of occasions that they co-occur with spoken backchannel forms, and the number that they are used alone (refer to the Appendix 4.4 for an extensive breakdown of these results).

| | Speaker | | TOTAL |
	<$M>	<$F>	
Nod, no Spoken Backchannel:	24	48	72
Nods with Spoken Backchannels:	22	22	44
	46	**70**	**116**

Figure 4.9: Frequency counts of backchanneling nods in the excerpt.

Figure 4.9 also demonstrates that there are a total of 116 backchanneling head nods in the excerpt, with both speakers using nods more frequently than spoken backchannels. However, it should be noted that there is a less marked difference in use for <$F> then for <$M>[47]. The data shows a reversal of that seen with the spoken backchannels, with <$F> nodding more frequently than <$M>. <$F> nods 70 times, amounting to 60% of the total, whereas <$M> only nods a total of 46 times, amounting to 40% of the total.

Whereas Figure 4.2 indicated that the most frequent speaker uses the highest net amount, i.e. the raw frequency, of spoken backchannels (although as a proportion of the total, this exists at a similar rate to <$F>), Figure 4.9 shows that the least vocally active participant uses far more backchanneling head nods than spoken forms. This suggests that during this excerpt <$F>, the supervisor, is adopting a more passive listener role than <$M> at this point in the conversation.

Figure 4.10 shows that the most common of these nod types, for both speakers, are types **A** and **C**[48] (please refer to section 4.2.3.1 for a definition of each head nod type).

Nod Type		Speaker		TOTAL
		<$M>	**<$F>**	
	A	15	31	46
	B	11	9	20
	C	13	26	39
	D	2	1	3
	E	5	3	8
		46	**70**	**116**

Figure 4.10: Total frequencies of backchanneling head nod types.

[47] This finding provides the stimulus for premise 6, see Chapter 5 for details.
[48] This finding provides the stimulus for premise 7, see Chapter 5 for details.

These were used in a total of 85 of the 116 head nods used in the case study data (i.e. 73% of the total). Both nod types **A** and **C** are nods with a short duration. Those with a longer duration, specifically types **D** and **E**, only accounted for 9% of the total. It should be noted that the frequency of use of types **A, B** and **C** are very similar for <$M> (accounting for 15, 11 and 13 occurrences respectively), whereas for <$F> there is a marked difference between the use of nods with a short duration (**A** and **C,** which have a combined frequency of 57), compared to those of a longer duration (**B, D** and **E,** which have a combined frequency of 13).

<$F> nods without concurrent spoken backchannels ('*gesture alone*' backchanneling nods, see Evans et al., 2001) twice as many times as <$M> (see Appendix 4.4 for a breakdown of the frequencies of different nod types and functions, as evidenced in the case study data). In addition, nods that are used without backchannels were most likely to be of type **A** or **C** for this speaker. This is also true of <$M>, although the frequencies of use for these behaviours are significantly less for this speaker, as detailed in Figure 4.11.

Nod Type		Speaker		TOTAL
		<$M>	<$F>	
	A	8	25	33
	B	7	6	13
	C	6	15	21
	D	1	0	1
	E	2	2	4
		24	48	72

Figure 4.11: The frequency counts of backchanneling nods occurring without spoken forms.

4.3.3.2. Nods with(out) concurrent spoken backchannels

Figure 4.12 provides a detailed breakdown of the different types of head nods that co-occurred with the spoken backchannels in the data. In comparison to Figure 4.11, this figure shows that backchanneling nods proved more likely to co-occur with spoken backchannels for both speakers than to be used alone, although the extent to which this is true differs across the speakers[49]. Indeed, in the vast majority of cases (aside from type **D** nods) the frequency with which these nods are used with concurrent spoken backchannels is far greater than the frequency with which they are used in isolation.

| | | Speaker | | TOTAL |
		<$M>	<$F>	
	A	7	6	13
	B	4	3	7
Nod Type	C	7	11	18
	D	1	1	2
	E	3	1	4
		22	**22**	**44**

Figure 4.12: The frequency counts of different backchanneling nod types co-occurring with spoken backchannels.

Figure 4.12 indicates that the type **A** nods were most likely to co-occur with spoken backchannels functioning as IR or CNV tokens (for both speakers, see Appendix 4.4 for further details). 8 of the 13 (31%) nods (4 from of each functional type) for type **A** were of this nature, consisting of 3 by <$M> (14% of the total number of spoken and non-verbal backchannels used by <$M>) and 5 by <$F> (23% of all instances). Type **B** nods proved most likely

---

[49] This finding provides the stimulus for premise 8, see Chapter 5 for details.

to co-occur with spoken backchannels functioning as CNV tokens. 4 of the 7 (57%) type **B** nods seen were of this nature, with 2 by <$M> (9% of total) and 2 by <$F> (9% of total). Type **D** nods proved to be just as likely to co-occur with CON as CNV tokens, and type **E** nods were as likely to co-occur with either CON or ER tokens, although the frequency of these types was relatively small (1 occurrence for each).

Overall, more than half of the small nods (64%) of short duration (types **A** and **C**) enacted by <$F> co-occur with spoken backchannels adopting an IR function (9 from 14), while half of the small nods of a longer duration (type **B**) co-occurred with the spoken backchannels adopting the CNV functions across both speakers. The type **C** nods co-occurred with CNVs and CON on 39% (7 from 18) and 50% of occasions (6 from 12), respectively. This co-occurrence was shown to be far more likely for <$F> than <$M>[50].

For 11 of the instances where <$F> used a type **C** nod (61% of the total), 100% co-occur with either a CNV token or a CON token. Whereas, for <$M>, only 6 of the total number of nods used were of type **C** (39%) and only 2 (22%) of these co-occur with either CNV tokens or CON backchannels. It is interesting to note that for all other types of nods no significant difference in the frequency of use exist, instead patterns of use are fairly consistent across the speakers.

---

[50] This finding provides the stimulus for premise 9, see Chapter 5 for details.

4.4. Overview

The analysis above highlights a number of issues which need to be taken into account when embarking on a corpus-based approach to the analysis of gesture-in-talk. Some interesting relationships have begun to emerge following these basic frequency-based investigations, which are as follows:

- Backchanneling head nods are used at the same rate or more frequently than spoken backchannels in conversation.

- In general, the most common *types* of head nods used in discourse are of a short duration and/or intensity. Intense, complex and multiple nods are less frequently used.

- Nods are used more frequently with concurrent spoken backchannels than alone. Similarly, spoken backchannels are used more frequently with concurrent nods than alone.

- More engaged forms of spoken backchannels, those situated at the bottom of O'Keeffe and Adolphs functional model, tend to co-occur with longer and/or complex head nod sequences (types **B**, **D** or **E**), whereas the less engaged and more simple lexical forms of spoken backchannels most frequently co-occur with shorter nods (types **A** and **C**).

These preliminary findings have provided an insight into some of the pragmatic properties of backchanneling head nods and their relationship with spoken forms of this phenomenon. These findings are reformulated and utilised as specific premises (numbers 6-9) for further investigation as part of

the extensive analyses conducted in Chapter 6 (note that premises 1-5 are based on previous findings of backchannel research, cited in the literature review in Chapter 2).

While the case study has identified some interesting observations regarding the use of language and gesture in a dyadic communicative context, at this stage, it has not been possible to provide a graded taxonomy of gesture types and functions and their direct relationships to language use, form and function. Therefore, the initial observations made are in need of further qualification from the exploration of a larger and more varied data set, with more speakers and so on, before more detailed assumptions are made regarding the nature of spoken and non-verbal backchanneling behaviour, and the relationship between them.

4.5. Summary

This chapter has provided an outline of a corpus-based approach to the analysis of new MM datasets for CL enquiry. It has proposed how patterns of language and gesture use can best be examined in records of communicative episodes, providing a blueprint for the main study analysis which is undertaken in Chapter 6.

The case study has illustrated that the proposed methods are both effective and appropriate for tackling MM data, as some interesting results and observations have already been identified as a result of this analysis. However, since this chapter has investigated only ten minutes of data, no definitive conclusions about patterns of backchanneling behaviour have, as

yet, been drawn-up as a consequence of this. Instead the focus has been on illustrating and testing the analytical approaches discussed within.

No real problems, beyond those discussed in section 4.2, were encountered when investigating this data, and while it is understood that the subsequent analysis of the five-hour dataset in Chapter 6 is necessarily more time-consuming, the fact that a formal framework for analysis has been developed here means that methodological problems and challenges faced in this analysis should be kept to a minimum. This matter is re-addressed accordingly as part of Chapter 6.

The approach used in this chapter is essentially analyst-led, utilising a system for detecting and encoding features which is manual, inferential and potentially very time consuming. Prior to the main five-hour study, it is perhaps appropriate to look towards more automated methods to facilitate the analysis of larger scale and more varied corpora. Chapter 5 examines this notion of 'automating the approach' in more detail. It provides a critical review of an intelligent digital gesture tracking algorithm designed to facilitate the investigation of forms of backchanneling nods in discourse, examining the practicality of this system and the relative advantages and disadvantages of using this method in preference to the manual approach investigated in the present chapter.

Chapter 5: Automated Analysis Techniques for Multi-Modal Corpora

5.1. Introduction

The principal function of this chapter is to take stock of the current state-of-the-art in MM CL methodology, and to propose some future technological developments in this field of research, based on the findings and observations made to date in this thesis. Overall the chapter intends to accomplish the following:

- Explore current, and developing, technologies that will enable the automatic detection and analysis of backchanneling head nods.
- Assess the ease and accuracy with which this is undertaken, discussing the technological and practical problems faced when using such methods.

Effectively, the chapter explores the potential for automating the processes of head nod detection, definition and codification as a means of allowing the analyst to search and manipulate large-scale MM datasets quickly and efficiently.

In order to achieve this, the chapter pilots a head tracking algorithm that has been designed to carry out such analyses; comparing the results from this with those already extracted by manual methods used in the case study in Chapter 4. By comparing the results, it will determine which method is

currently most adept at providing a proficient and systematic, multi-modal, approach to the analysis of 4th generation corpus data.

5.2. Automating the approach

5.2.1. The head tracker

Thus far the concern of this thesis has been with examining approaches to MM CL based research, through the implementation of a methodological strategy for the capture, analysis and re-use of video based corpus data. While it is true that the corpus technologies outlined so far are definitely novel (i.e. the hardware and software examined, such as the DRS and the NMMC), the approach to analysis as used in the case study is possibly less so. Instead, what is presented is a purely manual method, one which is referential; reliant on the skills of the analyst. Thus, the accuracy and functionality of this may be questioned when, for example, attempting to identify and encode a wider range and/or more complex forms of gesticulation.

In addition, while in reality the manual extraction and observation of individual head nods as they occur in ten minutes or even five-hours of video data is fairly unproblematic, it is logical to say that in 100 or even 10 hours of data it would very time-consuming and near impractical to administer this technique effectively. So, rather than relying on the ability of 'the trained analyst' to 'inspect and closely re-inspect gestures or tokens of interest and determine the most *appropriate* code for that item' (see 4.2.3.1 for further details), it is presumed that with the use of an automatic tracker a more definitive account of movement can be provided. Such a method is likely to

operate at high speed, helping to reduce the amount of time required to undertake such operations.

One such 'automatic' approach was built as part of the DReSS Project at the University of Nottingham, and tested as part of this PhD. This is in the form of a 2D head tracking device; the HeadTalk tracker (see Knight et al., 2006 and Evans and Naeem, 2007). This tracker uses a CV-based computational algorithm which is applied to the pre-recorded digitised video records, and subsequently reports, in each frame, the position of, for example, the speaker's mouth in relation to their eyes. An image of this tracker in operation is shown in Figure 5.1.

Figure 5.1: The HeadTalk tracker in action.

The circular nodes in Figure 5.1 are the tracking targets with a pre-defined granularity. The user is able to adjust the size of the tracked locations covered by these targets in relation to the size of the eyes and mouth of the participant, theoretically allowing users to track a range of images with different dimensions. This flexibility has proved particularly useful when using close-up images in which participants have larger eyes and mouths.

These targets are manually positioned at the start of the video and as the tracking is initiated a horizontal line is automatically drawn in the centre of these three nodes, marking an initial y-axis location with position 0. Subsequent vertical head movements are shown as causing a marked change in the y-axis in a + or − direction; with + being a head up movement and − being a head down movement. The horizontal line also rotates to the left and right depending on the position of the eyes, in order to monitor the 'angle of motion' of the head around the y-axis.
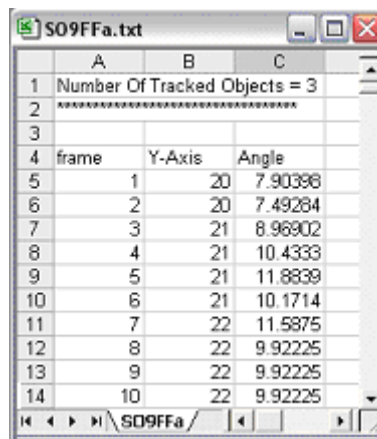
The observation of the head angle from one tracked frame to the next may prove invaluable to the analyst, as such information can help to identify the characteristics of specific types of head movement, for example, head shakes or head rotations as distinct from an up-down movement sequence associated with a head nod.

Finally, the HeadTalk tracker is also designed to allow the analyst to track more than one image in the same frame simultaneously, such as both speakers participating in a single supervision session.

5.2.2. Reading tracking 'outputs'

Again, this tracking algorithm determines the position of the y-axis, the resting or starting position of the head, and the angle of tilt of the head in each frame. This information is outputted as 'raw' data into an Excel spreadsheet, an example of which is seen in Figure 5.2.

Figure 5.2: An Excel output from the HeadTalk tracking algorithm.

It is currently the role of the human analyst to attempt to 'make sense' of this output. They are required to determine what sort of differences in y-axis location (as seen in column **B** of the output), over what range of angles (as seen in column **C** of the output), and over what time or frame span (as seen in column **A** of the output, 25 frames = 1 second) warrants definition as a 'nod' or 'no-nod' movement.

Given that this tracking system is still relatively new, the reliability and accuracy of the system requires to be tested. This is the function of the current chapter. Consequently, parameters for a 'nod' or 'no-nod' gesticulation

sequence are being continuously refined and re-defined during the further development of this tracker, i.e. this is an ongoing, iterative process.

5.3. Analysing the case study data using the head tracker

5.3.1. Data

To achieve the aims set forth in this chapter, it was decided that it was appropriate to utilise the case study extract as 'data' here. As, in terms of scalability, it is logical to test this innovative tracking facility on a small dataset, rather than attempting to negotiate the five-hours of video from the main study, as there is no guarantee of the cost-benefit of doing this. Again, the fact this 10-minute extract featured a large quantity of head nod movements, suggests that it would provide adequate stimuli for rigorously testing the tracker.

In effect, this mini-study functions as a secondary case study in this thesis. As with the previous case study, the observations and findings made as a result of this research will provide a strong case for investigating a more extensive dataset, although, in this case, such an investigation falls beyond the specific remit of the current study.

5.3.2. Approach

The following basic steps were taken in operating the tracker:

1- **Initiate tracker:** Locate facial regions on the tracker and run the software.

2- **Process output:** Map patterns in movement according to the Excel output, noting where, initially, dramatic changes in y-axis position occur across 25 frame intervals.

3- **Compare results:** Compare the tracking output with the manually prescribed codes to assess whether the automated and manual detection are congruous.

4- **Repeat process:** Track the complete ten-minute extract, comparing and assessing the accuracy of the results seen for each speaker, summarise the findings.

5.3.3. Results

In order to assess the relative accuracy of the tracker, that is, the proficiency of the system, a logical 'point of entry' into the data is to focus on specific movements that are detected as being significantly higher or lower than the mean y-axis position of the head. It is hypothesised that those y-axis movements which differ significantly from the mean are of a head-up or head-down nature, in other words a sequence of movements that potentially corresponds to a head nod.

To explore this hypothesis further, the tracking outputs generated from each speaker can be plotted graphically. Figure 5.3 maps the relative y-axis position of the head for <$M>, mapping the up-and-down motion of the head over time, denoted by the progression of frames of the video. A more detailed, 'raw', frame-by-frame breakdown of the head tracking results can be found in file 5.1 on the data disk for <M> (the left hand side of the table), and real-time

video records of the tracker in action for this speaker can also be found on the data disk (file 5.2).
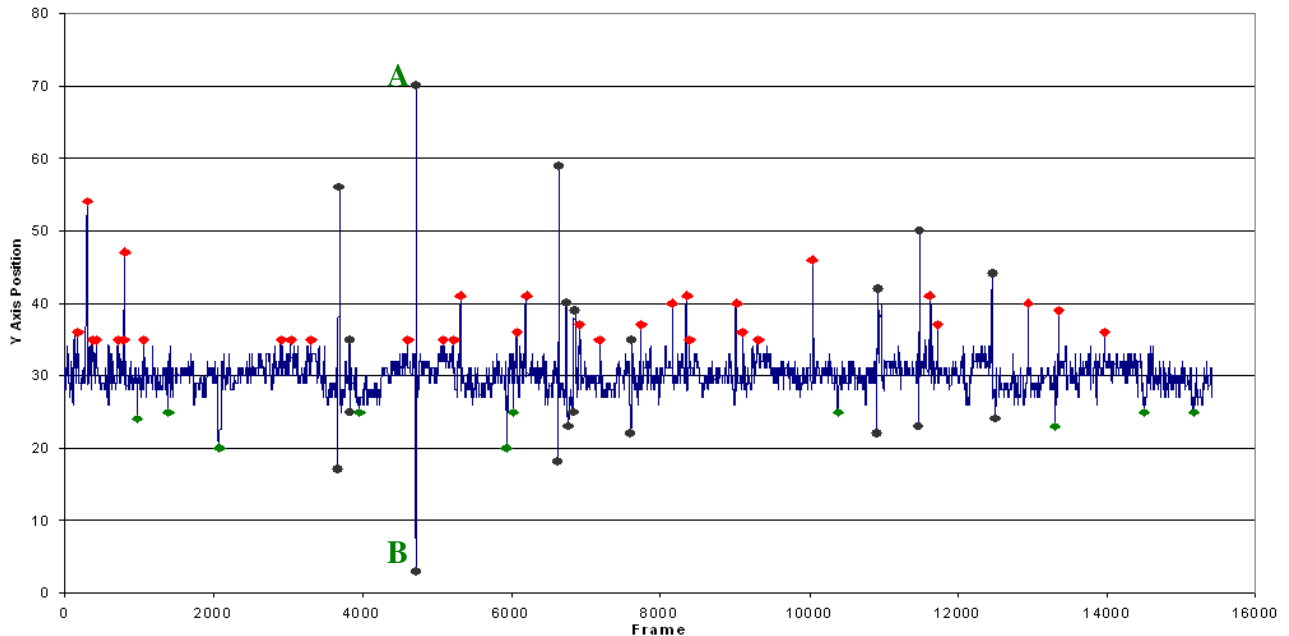
Figure 5.3: Tracking the head movements of <$M> throughout the ten-minute case study data.

Although a head nod is theoretically seen as an up-and-down movement, for the initial analyses of the tracked output it is appropriate to also explore situations where an intense up or intense down movement occur in isolation. These are situations which witness no preceding and/or following up or down movement of the head. This is because, as seen in Chapters 2 and 4, a nod does not necessarily always comprise of identical forms of movements in both directions. So, in instances where an intense movement is used, this may only be evident in one direction or the other.

In order to investigate the most intense head-up and head-down tracked movements of <$M>, those frames which have a y-axis position above and below circa 2 standard deviations (S.D.) of the mean head location can be examined in greater detail. This includes frames that have a y-axis output reading within the range of 25≤ y axis position ≥35 (refer to Appendix 5.1 for details of these specific frames), and are movements which are considerably above or below the average and/or 'no-movement' value (mean value = 30.135, 1*S.D.= 2.0223, 2*S.D.= circa 5.0446, rounded down to 5). It is hypothesised that 2 S.D. from the mean is an appropriate figure to test since such emphatic movements are more likely to be attributed to some form of gesticulation, but not necessarily a nod, rather than simply a shuffle or a fidget. Behaviours such as fidgeting and shuffling are instead assumed to cause more subtle differences in the head position than a head nod perhaps would.

Since the raw tracking output deals with a frame-by-frame account of the tracking results (see file 5.1 on the data disk), it is useful to group 'clusters' of frames that are located within this range of y-axis positions in order to make the analysis of the data more manageable. As a working benchmark, a 'cluster' is taken as a collection of up-or-down outputted movements that lie within the span of 25 frames of each other, and within the y-axis range given above. Therefore, these are groups of movements that are above or below 2 S.D. of the mean which exist within a 1 second time frame of each other. Although a 1 second margin appears slight, head nods can range extensively in terms of intensity and duration, which means that to best allow us to identify

a range of different head nod movements; from short to long, types **A** to **E** nods, such a small margin is required, possibly even smaller.

For <$M>, 599 frames, from a total of 15403, outputted y-axis coordinates within the 25≤ y axis position ≥35 range. These results can be clustered into a total of 40 intense head-up sequences (regarded as *peaks* hereafter; where the y-axis position is ≥35 for some or all of the frames across the cluster range), which are marked with red nodes in Figure 5.3. In addition, there is a total of 21 intense head-down clusters seen (regarded as *troughs* hereafter; where the y-axis position is 25≤ for some or all of the frames across the cluster range). These are marked with green nodes in Figure 5.3. Appendix 5.1 provides details of the tracking outputs for all intense peak and trough clusters seen in this video excerpt (see Table 1 of Appendix 5.1 for details on the most intense peaks, Table 2 for the most intense troughs), suggesting a range of movements that fluctuate between the y-axis positions of 3 to 70 (refer to 'min' and 'max' values in Appendix 5.1).

Table 3 of Appendix 5.1 provides a list of the ten clusters of frames where the head-up and head-down movements overlap or correspond to one another, within a 25 frame span. These are marked with black nodes in Figure 5.3 (refer to the 'max' and 'min' values detailed in Table 3 of Appendix 5.1 for specific y-axis values of these 10 clusters). These ten instances are, therefore, assumed to show where intense a head-up motion(s) is followed by a head-down motion(s), or in reverse; mirroring movements that we assume to be outputted in the case of emphatic head nodding behaviour.

The most emphatic peak type movement used by <$M>, as detected by the tracker, is marked as point A on Figure 5.3. This occurs between frames

4722 and 4732, ranging from 36 to 70 on the y-axis across this frame range.

This peak is immediately preceded by the most emphatic trough movement

seen in the data, as shown by point B on Figure 5.3. This trough, ranging from

3 to 25 on the y-axis values, exists between frames 4711 to 4721 on the

figure. This close succession of a peak and trough therefore exists as part of

the most pronounced up-and-down movement detected by the tracker,

marked as 'combined peak and trough 3' in Table 3 of Appendix 5.1. It is

relevant to note that this episode does in fact correspond to a backchanneling

head nod, as defined in Chapter 4. The conjuncture, at which this nod is used,

in context of the rest of the conversation, is shown in Figure 5.4 (refer to

Appendix 4.1 for a key to the coding used in this transcript; also refer to

Figure 4.1 in Chapter 4):

```
<$1> Yeah yeah <$X> that's | that is <\$X> still clever if it can do that.
<$2> <$=> Well I <\$=> <$=> I'm not quite sure <\$=> cos I've only read
the conclusion so far <$E> laughs <\$E>+
<$1> <$E> laughs <\$E>.
<$2> +But erm I'm not quite sure how they're tagging it for them or whether
they're just doing+
<$1> Right.
<$2> +you know key word searches+
<$1> Yeah yeah.
<$2> +I think they are just doing key word searches for them+
<$1> Right. Oh right yeah.
<$2> But.
<$1> Oh so you could set it to go look for like or as if or+
<$2> Yeah I think so.
```

Figure 5.4: Exploring the most 'intensive' nod from <$M> in the case study

data.

The section of transcript depicted in Figure 5.4 shows that a backchanneling

head nod of a long duration is enacted by <$M>. This starts at a mid-turn

point of speech by <$F> and continues while <$M> backchannels with the IR token *right,* and the subsequent CNV string *yeah yeah.* This specific nod was initially classified as being of type **E**, a 'multiple nod, comprising of a combination of types **A** and **B**, with a longer duration than types **A** and **C**' (see 4.2.3.1 for details).

Despite this initial success, such a manual-automatic detection agreement fails to exist at a constant rate for the remaining instances of 'combined peak and trough' sequences. In fact it is in a mere 2/10 (only 20%) of cases where the automatically tracked nods correspond to manually ascribed nods for the 10 intense clusters seen in Appendix 5.1.

This low success rate also holds true for the majority of the single head-up and head-down movements. From the 40 'intense' peaks detected by the tracker (see Appendix 5.1), 11 were manually pre-coded as non-verbal backchannels, with a further 1 as a non-backchanneling head nod (so 12/40, i.e. 30%) whereas, this figure of successful detection stands at only 1.5% for the intense troughs (3/20 instances), see Table 2 of Appendix 5.1 for further details.

In around half of the successfully tracked episodes, the nods that were detected were actually of types **C, D** and **E**; thus of the most intense types (5, 2 and 3 respectively). However, at this point it is relevant to note that multiple peaks and troughs may combine as part of an extended nod movement, see Appendix 5.1 for further details on nod numbers and codes. So of the total uses of these three nod types, as determined in Chapter 4 (see Figure 4.4), a mere 35% (6 different nods across the peaks and troughs) were correctly

identified as a consequence of the tracking analysis, although this includes 62% of head nod types **D** and **E** (5 of the 8 labelled in Chapter 4).

It should be noted that for the remaining instances where peak and trough movements were presumed to occur for <$M>, those which were not matched to nod movements as defined in Chapter 4, some other form of head and/or body movement *was* present. In the cases of p3,4,5 and t1,9,10 (for details on the specific frames that the 'p' (peak) and 't' (trough) codes represent, please see Tables 1 and 2 of Appendix 5.1), the speaker moves his whole body and head around as he is explaining a concept to <$F>, while in the cases of p34 and t15, the speaker moves his hand to his mouth, or scratches and moves his head around, again causing the tracker to lose its' target. Indeed across the ten-minute excerpt <$M> rarely sits still, instead he constantly fidgets and moves around in the chair. In such cases, the tracker is not inaccurate in suggesting that movement *is* occurring, it is just not always the type of head movements that are of interest.

Similar results to these are seen when looking at slightly less intense peak and trough movements across the <$M> tracking outputs. These movements are regarded as 'medium-sized' nods hereafter. Medium-sized peak and trough movements are defined as those within 1*S.D. of the mean y-axis position, so within the range of 26 ≤ y-axis position ≥ 34. Individual and clusters of frames that exist within this range are detailed in Appendix 5.2 (Table 1 for the medium-sized peaks, Table 2 for medium-sized troughs, and Table 3 for a combination of these). These are also plotted in Figure 5.5.

Medium-sized head peaks are denoted by red nodes in Figure 5.5 whilst the troughs are depicted as green nodes. Combined peak-and-trough

sequences, in other words those which exist within 25 frames of each other, are represented by black nodes in this figure.
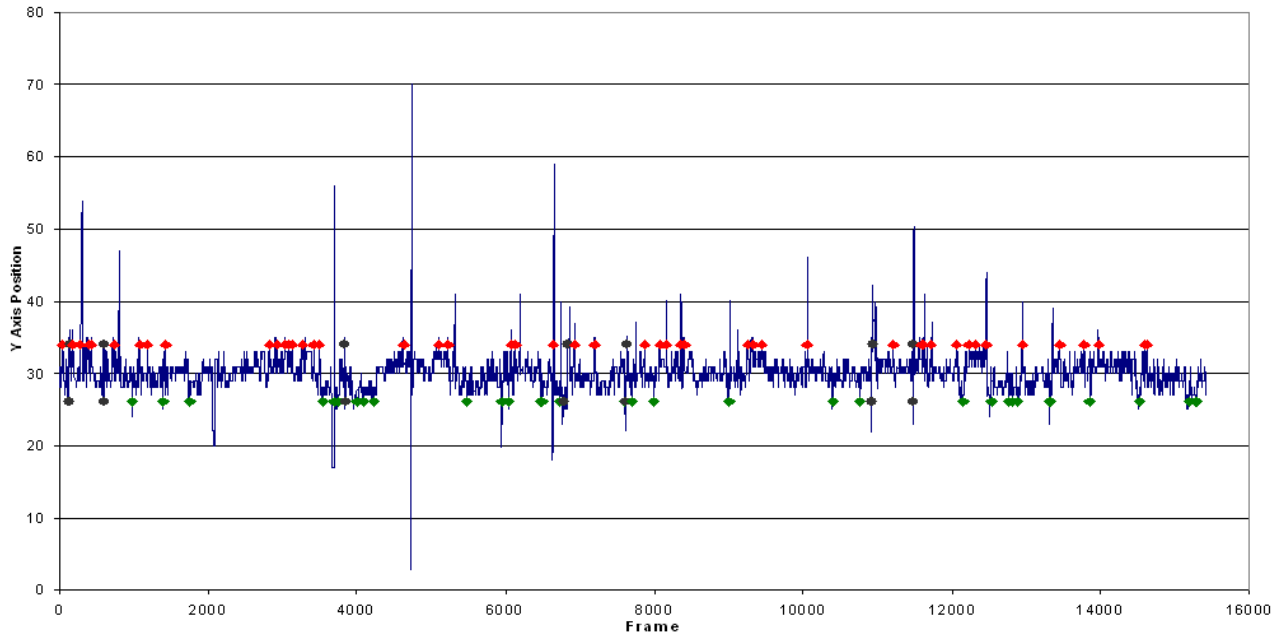
Figure 5.5: 'Medium' sized head movements enacted by <$M> throughout the ten-minute case study data.

Appendix 5.2 shows that, of the medium-sized movements, 58 peaks (labelled as 'sp', see Table 1) and 40 troughs (labelled as 'st', see table 2) were detected for <$M>, amounting to a total of 98 movements. Only 22 (19 different nods) of these movements were comparable to those previously defined, in Chapter 4, as comprising part or all of a head nod movement (15 for peaks and 7 for troughs). This amounts to a 22% success rate for correctly identifying head nods, which is the same rate seen with the analysis of the more intense head nods where the tracker detected 60 relevant peak and

trough movements, of which 14 corresponded to manually defined parts of nods; a total of 9 different nods. This is not a particularly impressive result.

What *is* more impressive is that across both the intense and 'medium-sized' nods (i.e. if we combine these results, as seen in respectively in Tables 1 and 2 of Appendix 5.2) performed by <$M> a total of 28 different head nods have been successfully detected (as identified in Chapter 4), in 36 one-second clusters of movement (across the individual peak and trough analyses).

If the first columns of both the peak and trough tables given in Appendix 5.3 (tables 1 and 2) are examined, it can be seen that this total of 36 is from a grand total of 97 (58 + 39) occasions in which the head was identified as moving up or down on the y-axis at a range of circa ≥1 S.D. of the mean y-axis value. This suggests that for 29% of the data explored, the tracker was successful in detecting the existence of head nod movement. The majority of which were labelled as backchanneling head nods, rather than other forms of nods.

The percentage of each nod types detected in this analysis, from the total of 28 *different* nods, is detailed in Figure 5.6. However, this table does not account for the additional movements that were automatically detected as nods which proved not to be, when compared with the manual analyses:

Nod type	Frequency detected by manual efforts	Frequency detected automatically	Automatic detection success rate (%)
A	14	11	79
B	11	7	64
C	13	5	38
D	2	1	50
E	6	4	67
	Total = 46	**Total = 28**	**Average = 61%**

Figure 5.6: Comparing automatic and manual methods of MM data analysis (<$M>).

It is now appropriate to consider the other participant; the female supervisee in order to assess the consistency of the results gained from the tracker; in comparison to those outputted for <$M>. When assessing the 'intense' movements enacted by this speaker, those frames that had a y-axis position within circa 2 S.D. of the mean (mean value = 25.33632, 1*S.D.= 2.247297, 2*S.D.= circa 4.49458, rounded up to 5) were again focused upon. This includes movements within the range of 20 ≤ y-axis position ≥ 30. Figure 5.7 charts the y-axis position of each tracked frame, marking the most intense peaks and troughs seen

Appendix 5.4 documents the clusters of individual intense *peak* and *trough* movements as well as clusters with a combination of peaks and troughs for this speaker, within a 25 frame range. A more detailed, 'raw', frame-by-frame breakdown of the head tracking results for this speaker can be found in file 5.3 on the data disk, and real-time video records of the tracker in action can be found in file 5.4 on the data disk.
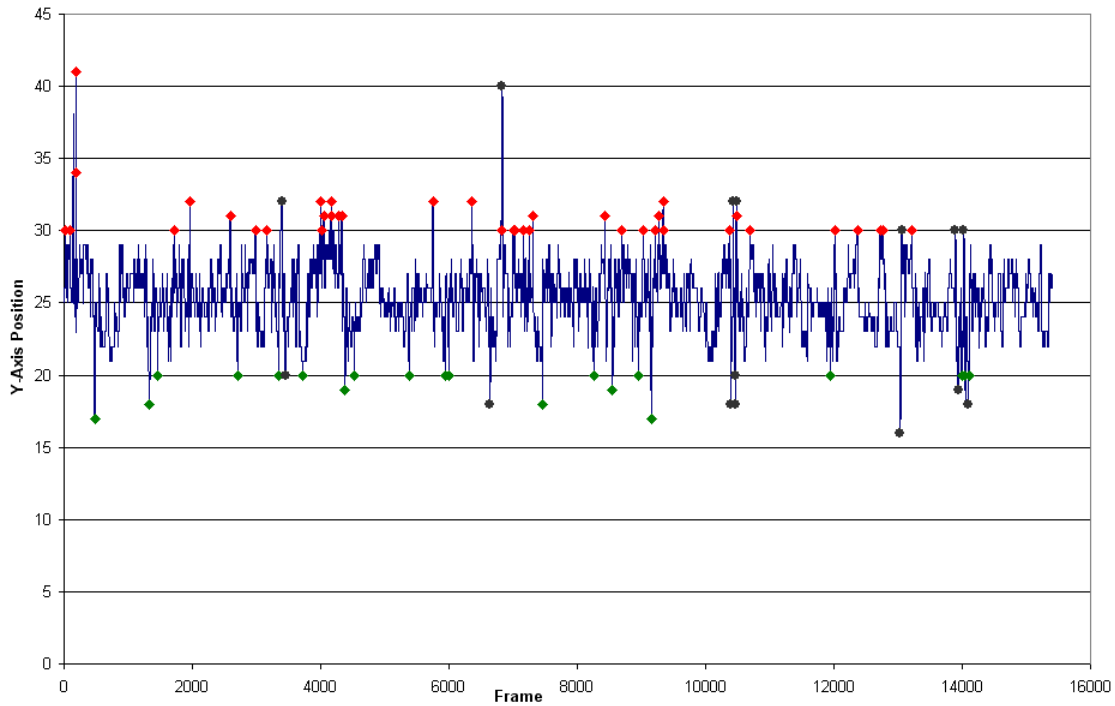
Figure 5.7: Tracking the head movements of <$F> throughout the ten-minute case study data.

In total, 42 intense peak (head-up; represented as red nodes in Figure 5.7, see also Table 1 of Appendix 5.4 for more detailed results) and 24 intense trough movements (head-down, represented as green nodes in Figure 5.7; see also Table 2 of Appendix 5.4) are seen in the case study sample for <$F>. Only 6 of these are instances where peaks and troughs co-occur within close vicinity of each other (as an up-down cluster; shown as a black node in Figure 5.7, also see Table 3 of Appendix 5.4).

On closer inspection of the original video excerpt, it was seen that in reality only 14 of these peaks (33%, from a total of 42) and 8 of the troughs (33%, from a total of 24) align to movements defined as head nods in the manual analysis. One other peak was defined as a head nod following this tracking and analysis process, but this was not defined as a non-verbal backchannel.

In addition, of the 6 intense head up-and-down clusters observed in this data, 4 (67%) were comparable with the head nod movements defined in Chapter 4. In short, for <$F>, although the tracker was able to detect and appropriately define the existence of some of the most intense head movements as nods, in the majority of cases it proved to be unsuccessful.

In terms of the other 'intense' movements depicted in Figure 5.7, the tracker incorrectly detected the following types of head movements as nods:

- Jerky movements when the participant was laughing, fidgeting in the chair, i.e. moving forward and backwards, as with t10, t16, p5, p9, p10, p14-18.

- Cases where <$F> scratched her head erratically or where a hand or arm obstructed the tracking target, as with t8, t9, p19, p42.

- The raising of the head when moving away from the paper, e.g. p21.

- Lifts and/or flicks of the head at the start of turns when the participant beans to explain a point and moved her body around for emphasis, as with t14, t15, t21, t24, p6, p8, p13, p23-28, p31, p37-8.

- Lowering of the head when the participant looked at the papers in her hand, as with t11, t12.

As with <$M>, it is obviously impossible to somehow attempt to inhibit the rate at which such fidgeting and shuffling or rotation of the head occurs since such movements are a characteristic facet of human behaviour. Attempts to control such behaviours would compromise the authenticity and naturalness

of the data. In reality, some speakers are likely to move around a lot more than others, depending on a wide range of discursive, personality-based (some are more 'active' than others), interpersonal (the relationship between speakers) and environmental traits (e.g. room temperature, or the level of comfort provided by the chair).

Further to the most 'intense' nods enacted by this speaker, it is again interesting to examine her slightly less emphatic head movements; the 'medium-sized' nods that display a y-axis value within 1*S.D. of the mean. These include frames with an output within the range of 21 ≤ y-axis position ≥ 29. The frames and frame clusters relevant to this range are detailed in Appendix 5.5, and plotted in Figure 5.8 (see Table 1 in Appendix 5.5 for medium-sized peaks, Table 2 for medium-sized troughs and Table 3 for a combination of medium-sized troughs and peaks).
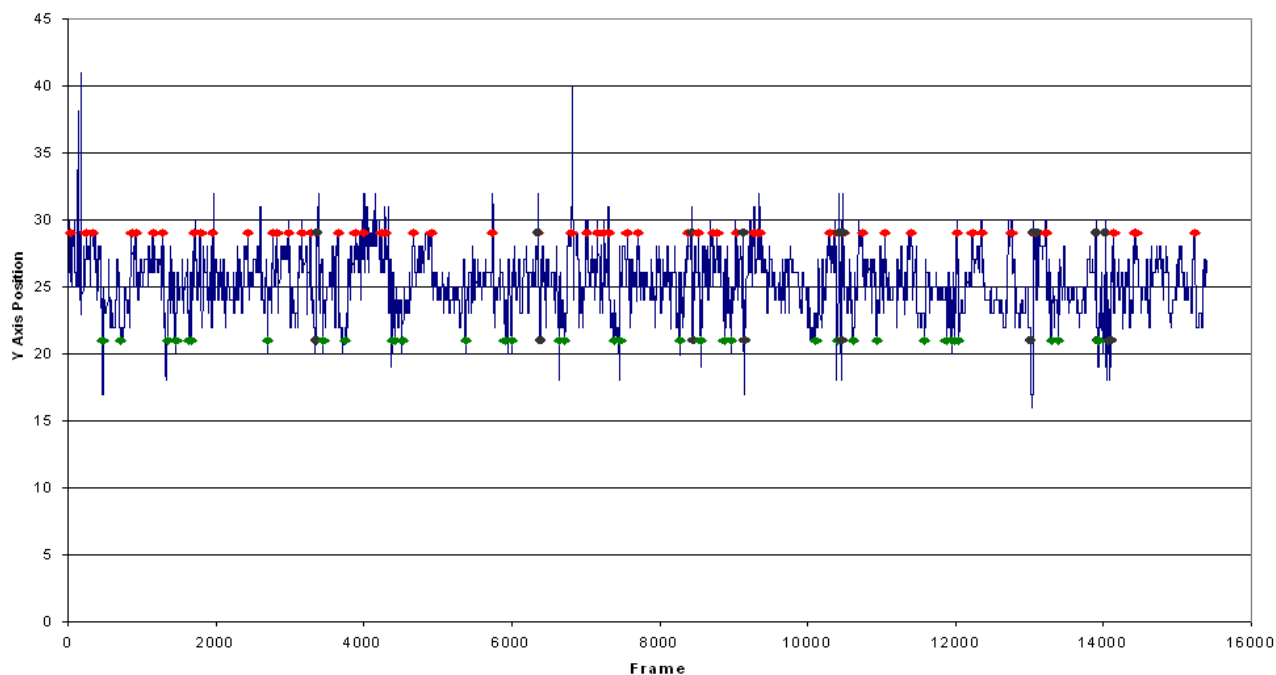
Figure 5.8: 'Medium' sized head movements enacted by <$F> throughout the ten-minute case study data.

Again, as with Figure 5.7, here the medium-sized head peaks are presented as red nodes, troughs are the green nodes and, finally, combined peak-and-trough sequences are denoted by black nodes (i.e. those which exist within 25 frames of each other). Overall, the tracker detected a total of 106 peaks and troughs within this y-axis movement region, 34 of which corresponded to the manually pre-coded nods. This amounted to 31 different nods, including 3 trough-peak combinations. This suggests a 32% accuracy rate for the tracker, which is a similar rate to that generated by tracking <$M>.

When considering the intense and medium-sized nods together (see table 3 of Appendix 5.6, see also Tables 1 and 2 respectively for combined intense and medium-sized peaks and troughs), it is noticeable that in total 38 different nods were successfully detected by the tracker on 57 occasions (i.e. over 57 one second clusters). In other words, 67% of movements existing within ≥1 S.D. of the mean y-axis value correspond to backchanneling head nod movements. This is from a total of 103; 57 from the total of 64 peaks + 39 troughs (see column 1 in Appendix 5.6, Tables 1 and 2, for further information). This appears to be quite an impressive success rate, especially in comparison to the 29% success rate seen with <$M>. Figure 5.9 indicates, based on these 38 different head nods, the specific types of nod that were successfully detected.

Figure 5.9 shows that although it might be expected that the most intense nods, and those of a medium-sized intensity, would correspond to nod types **C, D** or **E**, in this situation it is actually nod types **B** and **D,** those with a long duration, that are most successfully tracked. In both of these cases 100% of

the nods are successfully detected using automated methods, as accounted for in Figures 4.11 and 4.12 in Chapter 4.

Nod type	Frequency detected by manual efforts	Frequency detected automatically	Automatic detection success rate (%)
A	32	13	41
B	9	9	100
C	26	14	54
D	1	1	100
E	2	1	50
Total = 70	**Total = 38**	**Average = 54**	

Figure 5.9: Comparing automatic and manual methods of multi-modal data analysis (<$F>).

However, it is important to note that these nods actually only comprise *circa* a quarter of the overall amount originally detected, so in other words the overall frequency with which such nods were used by this speaker proved to be fairly low. Again, these figures are particularly impressive when compared to <$M>, who generally fidgeted much more during the ten-minute excerpt, causing the tracker to lose the targets more frequently than for <$F>.

5.4. The functionality of the tracker

5.4.1. Different speakers and videos

Further to the results obtained from the tracking output, it is also necessary to discuss some other basic points of the tracking system's functionality; namely the practicality of using such a system to interrogate large-scale linguistic

datasets. This includes a consideration of user-based requirements for using this tracker and other similar systems, such as the technological resources and nature and quality of the data available for analysis.

At this point, It is important to note that the tracker is actually able to track using two different 'filters' (i.e. two versions of the same tracker), which determine the speed at which the tracker is operated. These are the KAMS filter (the Kernel Based Annealed Mean Shift Algorithm; see Knight et al., 2006 and Evans and Naeem, 2007) and the Mean Shift tracker, the latter being the algorithm that has been used with the data thus far. It was discovered the KAMS filter generally took up to 25 minutes to track 30 seconds of data, making the process very lengthy and impractical to use when analysing large quantities of data.

Working in real time, the Mean Shift filter proved to be considerably quicker, allowing the user to track larger volumes of data at speed. However, this was not necessarily always the more accurate of the two. At times throughout the tracking of data, the eyes and mouth targets are frequently 'lost' by the tracker and while both filters may be stopped in-action and re-administered to the correct position, the speed of the Mean Shift can in theory result in a longer frame delay between the analyst detecting the loss and 'debugging' the system (debugging is a process that involves the constant redefinition and relocation of tracking targets). This is owing to the accelerated tracking speed of this filter, i.e. it is likely that a longer span of frames is affected by the 'loss' when using the Mean Shift filter. Despite this, after preliminary tests of the Mean Shift tracker, it was found that when the targets

are lost and have to be relocated, it is only in rare instances that the problem of the 'frame delay', mentioned above, actually occurs.

Should the analyst be required to re-administer the slower KAMS filter over sections of video where 'losses' frequently occur (when using the Mean Shift filter), in order to reduce the amount incurred and the frequency with which the 'frame delay' problem is faced, this process is likely to still be quicker than using the KAMS filter throughout.

As a further point of discussion, it should be noted that the tracker is at its most effective when applied to a high quality video (.avi), with a high resolution. In other words, the tracker is most adept at processing videos where the image of each participant is close-up and as large scale as possible. Smaller, lower quality images are more likely to lose the tracking target locations instantly. This requirement proved to be slightly problematic when dealing with the streamed two-party videos from the supervision sessions, as the standard size of these tends to be relatively small, especially once compressed for use in the NMMC corpus. This reduction in the size and associated quality of the images caused the tracker to readily lose the target locations, making it difficult for the CV algorithm to adequately track movements. In consequence, it proved more practical and accurate to utilise the original, un-streamed videos of each single participants during this tracking process, rather than these smaller streamed images.

However, by using the individual source videos, rather than those which have been aligned, the process of tracking obviously becomes a much lengthier one. This is because the image of each individual participant requires to be tracked in turn, rather than simultaneously. Nevertheless, given

197

that the synchronised split screen videos require more debugging, which in itself is very lengthy process, tracking the individual videos actually proves to be less time-consuming to undertake overall, making it more cost-effective in the long run.

Related to the problem is the fact that since .avi files are of a very large size, around 10 gigabytes for a 50 minute video (making them difficult to store in vast quantities on a standard PC), a phenomenal amount of processing power is required to run the tracker effectively without crashing the system. Indeed, whilst testing the tracker it was discovered that any stretch of video over 30 seconds long would cause this failure to occur (also partly attributed to a memory leak in the tracking algorithm), which was far from an ideal situation.

To overcome this, when undertaking some initial analyses of the data, the individual videos were cut into smaller components, amounting to 30 second clips, *a priori* to running the tracker. In effect, 40 different clips existed for the case study extract. Although processing 40 single clips is not overly time-consuming, with a larger dataset this technique becomes more difficult to manage. Consequently, the bigger the dataset, the more time is required to firstly construct these clips, then run the tracker and reorganise results.

Furthermore, although this segmentation made the tracker more functional, it questioned the reliability of the tracking output generated when adopting such measures. This is because it is naturally difficult to ensure that there is a consistency in the repositioning of the tracker from one clip to the next when the tracking targets are continually, and manually, re-located and redefined.

A further limitation of the tracking system is its relative unsuitability for use on other forms of real-life communicative data, aside from that used here. In different environments, perhaps with multiple speakers or in instances where recording is less static and/or 'on-the-move' (see 3.3.2.1 for further discussions related to this matter), associated problems with the clarity and closeness of the image, or extreme movement by participants, can cause the tracker to fail and/or generate inaccurate results.

To illustrate this limitation, as an extension to the main tracking case study, the HeadTalk algorithm was briefly tested on segments of lecture data taken from the NMMC, using the same approach to tracking as identified in section 5.3.2, above. The basic set-up for this recording included a single static camera positioned in front of the speaker (at a variable distance, from lecture to lecture), which was controlled by the researcher; adjusting the focus of the camera as required, when the lecturer moved around the room. Although, in reference to the discussions seen in Chapter 4, this data is not necessarily 'naturalistic', the images obtained present a range of challenges that are valid and similar to those that are likely to be faced when other types of corpus video data are used with the tracker.

Figure 5.10 outlines some of the basic problems faced when using this tracker for alternative datasets, using stills from the lecture component of the NMMC. When testing the tracker on the NMMC lecture data, it was discovered that in each instance, as soon as the recordings commenced, participants walked around, turned away from cameras (see images A and B in Figure 5.10), turned down/ off lights (see images C and D in Figure 5.10)

and/or started to hide behind objects including paper and/or lecterns (see images E and F in Figure 5.10).

Figure 5.10: Highlighting data reusability problems.

These behaviours meant that it was difficult, almost impossible, to use the tracking algorithm on such data, thus limiting the proficiency of the tracker on such datasets. Indeed it is difficult to even observe patterns of NVB or NVC in a manual way with this data, because it is obviously difficult to adequately

monitor behaviours which are hidden and/or out of view from the cameras; that is, behaviours which are not freely observable.

Finally, it is relevant to note that the tracker is only adept with dealing with image data in 2 dimensions. It does not process audio or textual information, so exists merely as a bolt-on functionality to a MM corpus tool-bench. This means that in itself the application is not technically multi-modal, however, since the present thesis has discussed its use in context of other tools, this is not strictly a limitation. Related to this, although the tracker may facilitate the encoding of movement, it does not enable the automatic prescription of *meaning* to forms of NVC. It is unable to re-contextualise the movement in the way an analyst would, such as perhaps determine whether a particular nod is a polar response or, indeed, a form of backchanneling behaviour. At present no tracking device is fully equipped for doing this. As identified in Chapter 2, the exploration of gesture-in-talk is necessarily centred on patterns of meaning, so it is vital to emphasise this limitation; the role of the human analyst thus remains central to this particular stage of the process.

5.4.2. Manual Vs automated head nod tracking and definition

In terms of cost-benefit, it seems that at present the use of the HeadTalk tracker for the automated analysis of MM corpus data complicates rather than simplifies the process of head nod detection and definition. Although, in general, the tracking algorithm was adequate at defining the points in the data where head and upper torso movements occur, these rarely proved to be instances of head nod behaviour. The tracker can be used to define movement from non-movement, however, the analyst is still required to

manually inspect every instance in order to check whether they are, in the case of this thesis, backchanneling head-nods or not. The analyst must sort the NVC from the NVB, then subcategorise them according to meaning and discursive function. This process, when combined with the initial time taken to even generate the raw results from the tracker, would again take the analyst a longer time than it would to undertake the entire analysis using more manual methods.

In short, if in the future the tracker had the ability to operate over large amounts of varied forms of video data at real-time speed (or even faster), and was capable of exploring a range of different forms of gesticulation with minimal target losses, it would be beneficial to use this method in MM CL exploration. However, at present the technological and practical problems associated with the tracker suggest that it is appropriate to use a more manual method for head-nod analysis until the tracking algorithm has been improved.

As a final note, it is possible to suggest that many of the problems associated with this system may perhaps have already been overcome by other, similar tracking devices. However, the difficulty with such a claim is no tracking system exists that is freely available to use, easy to access or operate, nor one which has been specifically tried and tested for use in MM CL research. Many current CV tracking algorithms are bespoke; designed for use on specific forms of data, such as sign language data, for example, (see Ong and Ranganath 2005) and/or for examining specific episodes of gesticulation. Such trackers lack utility beyond these requirements; aligning them with the key limitations listed in section 5.4.1 above. For examples of

alternative trackers see Isard and Blake, 1998; Morimoto et al., 1998; Pittner and Kamarth, 1999; Deutscher et al., 2000; Kawato and Ohya, 2000; La Cascia et al., 2000; Davis and Vaks, 2001; Colombo et al., 2003; Comaniciu et al., 2003; El Kaliouby and Robinson, 2004 and Deniz et al., 2004. Consequently, the HeadTalk tracker presently exists as the best system for undertaking the enquiries discussed within this thesis, and for this reason it has been used here.

5.5. Summary

This chapter has provided an extension to the research conducted thus far. It has concentrated on the second aim of the thesis; developing new approaches to the actual analysis of MM corpus data. The chapter has operated on the observation that when investigating gesture in human interaction, 'it is quite difficult to isolate…the single movements on the recording data and analyse them in detail' when using manual methods alone (Cerrato and Skhiri, 2003: 252). This problem is compounded when large datasets are utilised; for which a corpus-based approach ultimately intends to allow. It is this utility that sets this approach apart from mere video analysis methods. Although, at present, the HeadTalk tracker is not as accurate or efficient as desired, future developments in this area of research will help to improve the system. Therefore, at present the manual approach outlined in Chapter 4 is a more practical and appropriate method of analysis for the remainder of this thesis. Consequently the in-depth analysis of the five-hour sub-corpus carried out in Chapter 6 is undertaken predominantly using the more manually driven techniques and strategies outlined in Chapter 4.

Chapter 6: Analysing Backchannels in a Five-hour Multi-Modal Corpus

6.1. Introduction

This chapter provides an in-depth analysis of the patterns of backchannel usage in real-life conversation, as evidenced by a five-hour sub-corpus of NMMC data. While chapters 4 and 5 considered the case study, which was designed to develop and test a CL methodological approach relevant for MM data, the present chapter will discuss how this approach was utilised to examine whether any specific patterns and/or relationships were seen to exist between the use of spoken and non-verbal backchannels in the corpus.

This analysis is undertaken by the investigation of three key 'variables'; backchannel frequency, form and function, as follows:

- **Frequency**: The total number of words spoken by a participant and the total number of spoken backchannels and/or backchanneling nods they use, over the course of a conversation.
- **Type**: The patterns and characteristics of spoken and non-verbal backchannel use, focusing on the frequency of use of:
    - Particular *forms* and *functions* (i.e. CON, CNV, ER and IR) of spoken backchannels, and specific forms adopting a given function.
    - Individual backchanneling nod types (**A**, **B**, **C**, **D** and **E**).

- Patterns of ***co-occurrence,*** i.e. the simultaneous use of spoken and non-verbal backchannels. Questioning

    - Which nod types are most frequently used with spoken forms and/or functions.

    - Whether speakers more or less likely to use backchanneling nods with a spoken variety adopting a a) CON, b) CNV, c) ER or d) IR function.

These characteristics are investigated in relation to ten specific premises, which have been constructed with reference to both previous research detailed in Chapter 2, and the preliminary findings seen in Chapter 4. Each acts as a specific research question and are systematically presented, investigated and summarised in each of the sub-sections of 6.3, below.

At this point, it should be noted that the current chapter provides the initial analyses of the dataset, whereas the following chapter gives an in-depth assessment and a discussion of the relevance of these results. Thus, Chapter 7 will provide a linguistic commentary, discussing the possible reasons for patterns that emerge.

6.2. Overview of approach

6.2.1. Approach

The approach used for the analysis of this data is an adaptation of that used in the case study. This is illustrated in Figure 6.1.

Step	Variable	Description	Spoken					Non-Verbal				
1	Freq.	Raw counts of backchanneling phenomena (per speaker/ video)										
2	Form	Discursive/ movement structure	Lexical structure and whether it is comprised of a single token/ phrase					Intensity and duration of the nod				
			Simple	Double	Complex			A	B	C	D	E
3	Function	Linguistic function	Discursive function					?				
			CON	CNV	ER		IR					

Figure 6.1: A matrix for the annotation and analysis of backchannels in discourse.

It is necessary to note the '**?**' section seen in the bottom right corner of this figure. This indicates that while various past studies have noted, for example, that backchanneling nods function in a variety of ways in discourse and/or share many pragmatic similarities to spoken forms of behaviour (as discussed in Section 2.4 of Chapter 2, see Maynard, 1987; Maynard, 1997; McClave, 2000; Kapoor and Picard, 2001:1 and Cerrato and Skhiri, 2003), there is, however, currently no formal linguistic-based system that adequately classifies the discursive function of these elements. Furthermore, no classification system exists which provides descriptors of how such behaviours interact with spoken features in order to generate meaning in discourse. Consequently, following the analyses undertaken in this chapter, Chapter 7 will consider how best to fill in this '**?**' box, providing an insight into the specific roles and functions these behaviours adopt in dyadic conversation.

To investigate the variables, five stages of analysis are undertaken, as follows:

Stage 1: Specific non-verbal/ spoken behaviours are identified as backchannels.

Stage 2: The type of nod, linguistic form and discursive function of the non-verbal and spoken backchannels, respectively, are classified.

Stage 3: The frequency of each backchannel form, type and function are noted for every speaker in each supervision.

Stage 4: Instances of spoken and non-verbal backchannel co-occurrence are marked and frequencies of their associated forms and discursive functions, where known, are again noted for each speaker and across each supervision video.

Stage 5: Frequencies of spoken and non-verbal backchannels (co)occurrence from each speaker and/or video are compared and observations regarding interesting patterns made.

6.2.2. Data sample and labels

The corpus on which the following study is based comprises five hours (307 minutes) of video recordings which include 12 different speakers taken from six different one-to-one supervisory meetings featured within the NMMC. These meetings have all been transcribed and anonymised. For ease of reference, each supervision video is labelled according to the standard NMMC scheme as shown in Figure 6.2.
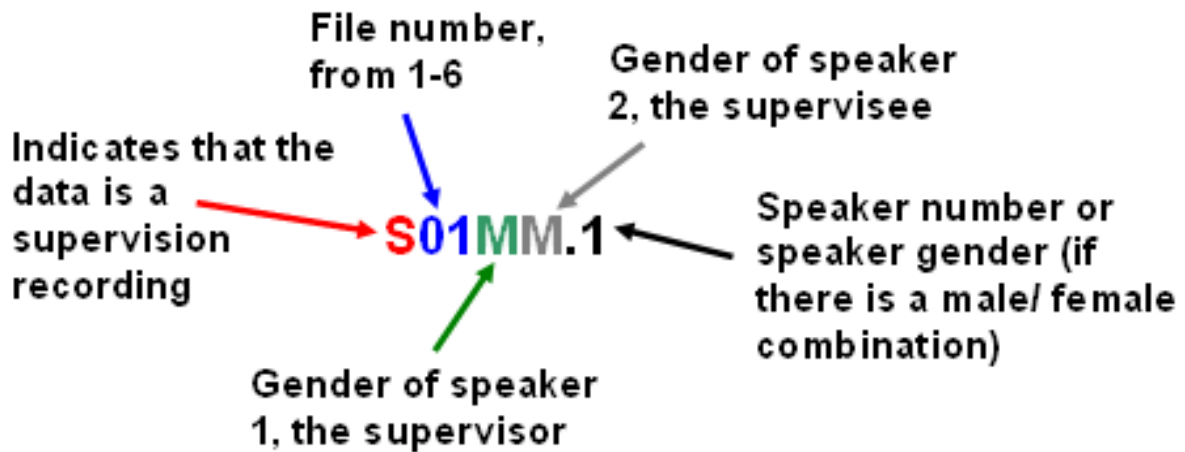
Figure 6.2: A guide to data labels used in the main study.

Each transcript has been marked-up using the same conventions used for the case study in Chapter 4, as seen in Figure 4.1 of the chapter. In other words, each spoken backchannel is highlighted, numbered and categorised according to the pragmatic functions discussed previously, and all instances where backchanneling nods occur have been underlined and numbered. This basic method allows the free observation of where such behaviours occur within and across turn boundaries. The annotated transcripts for the entire five-hour dataset, which include records of spoken and non-verbal backchanneling behaviour and running frequencies of these, can be found in files 6.1 to 6.6 on the data disk. These files are for supervisions S01FM to S06FF inclusive.

In order to draw observations from the data annotated in these transcripts, all instances where spoken and non-verbal backchannels occur and co-occur, along with their type and discourse function, where applicable, have been counted and collated into frequency-based tables. These can be found in Appendices 6.1 to 6.6 for supervisions S01FM – S06FF inclusive.

The total word count across these recordings is 56214 words without speaker or formatting tags. This total is distributed across each of the supervisions and individual speakers according to the values seen in Figure 6.3.

Speaker	Number of Words	Word Total	Total Time (Minutes)
S01FM.F	3754	8213	35.30
S01FM.M	4459		
S02MM.1	5266	8768	60
S02MM.2	3502		
S03MF.M	5834	8410	41
S03MF.F	2576		
S04MM.1	5066	8676	58.30
S04MM.2	3610		
S05MM.1	4306	11338	64
S05MM.2	7032		
S06FF.1	8154	10828	48
S06FF.2	2674		
		56214	**307**

Figure 6.3: The raw word frequencies of the main study data.

The justification for selecting this five-hour dataset was based on a combination of factors, namely the availability of resources and the amount of time available to conduct this research. The NMMC was being developed during the course of this study, so a bank of 125,000 words of videoed supervision data was freely available for use. It was decided that it would be more beneficial to take complete conversational episodes from the corpus, from 6 different dyads of speakers, rather than random extracts of talk.

The reason for this was to provide a better picture of the use of backchannels from the start to the end of each conversation, allowing for the examination of clusters of behaviour over time and the structure of discourse. Furthermore, only data from supervisory meetings from the same academic department were selected for analysis. This was in order to create a consistent recording context for the discourse episodes.

Given that the behaviours of 12 different speakers are analysed, this study can track not only simple characteristics of language and backchannel use demonstrated by a single speaker, i.e. idolectic properties, but also summarise patterns that exist across all speakers in each conversation. Consequently, the chapter initially discusses comparisons in backchannel use between the two speakers in each conversation, before the entire dataset is compared and more detailed analyses are made.

6.2.3. Statistical relevance testing

Before proceeding, it is relevant to discuss whether or not it is appropriate to integrate statistical relevance testing methods into the analyses undertaken here.

As Huberty explains, statistical testing has existed in some form for the last 300 years and is now commonplace in social science research (1993). Statistical tests such as the mutual information test (MI), log-likelihood, t-score, z-score and chi-square are extensively used for empirical linguistic study. Such tests, arguably the most common of which is the MI score, allow analysts to explore the relationship between two populations; x and y. They function to establish the probability of whether specific patterns or significant

relationships observed between the populations are likely to exist by mere chance (see Church and Hanks, 1990 for other examples of statistical relevance tests). These measures allow researchers to either strengthen an argument put forward following the analysis of data, or to support or contradict a hypothesis of a given study, with a statistically proposed level of 'confidence'.

In terms of CL research, statistical tests are normally used in studies of comparative corpora, when conducting large-scale statically-based comparisons of a given corpus to other, reference corpora. Despite this, there remains no common consensus regarding 'best practice' for the use of statistical tests in CL methodologies and indeed it is widely debated whether these measures should or should not be used. So while some CL researchers use these tests in their research, others avoid them entirely. Indeed, as Pedhazur and Schmelkin state (1991: 198) 'probably few methodological issues have generated as much controversy among socio-behavioural scientists as the use of [statistical significance] tests'.

Statistical testing essentially relies on the notion of randomness, of stimuli co-existing through chance. This element of randomness is at the crux of criticisms for using statistical testing in CL research. Sinclair questions 'why use chance as a criterion of relevance' (2008: 29) in CL analysis when, as Kilgarriff comments 'language users never choose words randomly, and language is essentially non-random' (2005: 263). So, in essence, when conducting CL based research, what is presented 'in front of us are not probabilities, but actualities, and those should be the focus of our attention' (Sinclair, 2008: 24).

This leads to the supposition that the essential problem with integrating statistical tests into a general CL approaches lies in the fact that these tests are often 'misapplied' (Daniel, 1998: 27, also see Halliday, 1992; Dunning, 1993; Stubbs, 1994; Kilgarriff, 2005 and Sinclair, 2008).

In effect, corpus linguists are dealing with records of real-life episodes, of language not numbers, and since 'there is no clear theory of how the frequency of linguistic features contributes to the meaning of individual texts' (Stubbs, 1994: 217), the use of statistical verification perhaps complicates corpus-based enquiries rather than simplifying this already complex approach to language study.

This criticism does not necessarily suggest that statistical testing should be disregarded completely within the field of linguistics, or in other social science disciplines, because such tests have traditionally proven invaluable to other areas of research. However, it does suggest that caution should be exercised when implementing such tests on corpora. Sinclair contends that CL methodology 'needs its own methods of statistical analysis, which should be purely descriptive and which should qualify linguistic concepts and categories' (Sinclair, 2008: 30); a point that is particularly relevant in light of the onset of *developing* MM datasets. Nevertheless, such methods have yet to be realised and thus 'the numerical and statistical side [in corpus linguistics] has scarcely begun' (Sinclair, 2005).

Given the above comments, this thesis adopts the position that there is no added value in using methods of statistical testing in the extended analysis presented in this chapter. This is because the present study is explorative, thus, as is often in the case for CL research, it seeks to 'describe and explain

the observed phenomena' (Sinclair, 2008: 30). This is in order to assist in developing an enhanced understanding of something which, at present, is not fully understood, i.e. the ways in which spoken and non-verbal backchannels interact to create meaning in discourse. It does not set about proposing or validating specific rules, nor does it aim to provide definitive conclusions about these behaviours, as this is practically impossible when dealing with human behaviour. Rather, it functions to provide an exploration of the validity of the questions and hypothesis outlined in Chapter 1 of the thesis. In other words, since the key aim of this study is to enhance our understanding of meaning in discourse, 'relevance' is seen in the ways in which such meaning is constructed, and not in a score provided by a statistical relevance test.

Despite this, in Section 6.3 of this chapter various numerical comparisons between datasets and speakers are made, based on the interrogation of the raw frequencies of behaviours, as seen in Chapter 4. However, these merely exist as simple percentage-based observations of the data, used purely to illustrate whether the patterns of behaviour seen for one speaker or video are similar to those shown by other speakers, or across the entire corpus. This technique operates as a point-of-entry into the data, working on the premise that is 'sufficient simply to count and list items' for CL analyses to be undertaken (Stubbs, 1996: 5); thus, it is appropriate for simple percentage comparisons alone to be made.

6.3. Results

6.3.1. Frequency of backchanneling usage

6.3.1.1. General observations

The first line of enquiry for this analysis is an exploration of the basic relationship between the *number* of words spoken by a participant and the *rate* at which s/he uses spoken backchannels and backchanneling head nods, from the total number of backchannels used by a particular speaker. This initial line of investigation examines the validity of the following premises:

1- 'Backchanneling occurs more or less constantly during conversations in all languages and settings' (Rost, 2002: 52, also Oreström, 1983; Gardner, 1998).

2- If one speaker dominates the conversation significantly then the other will backchannel more.

These premises suggest that the frequency of spoken and non-verbal backchannels will be fairly high across individual speakers and conversational episodes, amounting to a high rate of occurrence of these behaviours across the entire corpus. Having said this, it is expected that there is not necessarily a consistency in the frequency across both speakers in a given supervisory dyad at one given time over the course of a conversation. This is because the roles of the participants are likely to change throughout. So if at a given point one participant is acting predominantly as the speaker, the information giver while the second participant adopts a more passive information receiver role, (i.e. the listener), then backchanneling is likely to be used more often by the

second participant as this type of behaviour is more related to listening. What remains to be seen is whether there is a difference between the frequencies with which those who are assuming predominantly listener-based roles perform backchannels across the spoken and non-verbal mediums. Thus, whether a participant who uses spoken backchannels *x* number of times less or more often than the other participant is more or less likely to also use more non-verbal backchanneling nods, and at what rate of difference.

To test these premises, the raw frequencies of spoken and non-verbal backchannel usage, and word-use frequencies for each speaker and video in the corpus sample have been collated and are presented in Figure 6.4 (note: in this figures BCs = backchannels). Detailed breakdowns of these frequency counts are shown in sections 1, 2 and 5 of Appendices 6.1 to 6.6, for supervisions S01FM to S06FF inclusive.

Speaker	Words Count		Spoken BCs		Nods	
	Speaker	Total	Speaker	Total	Speaker	Total
S01FM.F	3754	8213	250	287	211	311
S01FM.M	4459		37		100	
S02MM.1	5266	8768	70	539	165	515
S02MM.2	3502		469		350	
S03MF.M	5834	8410	160	292	154	465
S03MF.F	2576		132		311	
S04MM.1	5066	8676	283	533	201	453
S04MM.2	3610		250		252	
S05MM.1	4306	11338	342	487	286	473
S05MM.2	7032		145		187	
S06FF.1	8154	10828	105	292	121	468
S06FF.2	2674		187		347	
		56214		2430		2968

Figure 6.4: The frequencies of spoken backchannels, non-verbal backchannels and words across each speaker and video in the five-hour corpus.

Building on this, Figure 6.5 illustrates the ratio-based relationships between spoken/non-verbal backchannels and word usage for each speaker in this corpus.
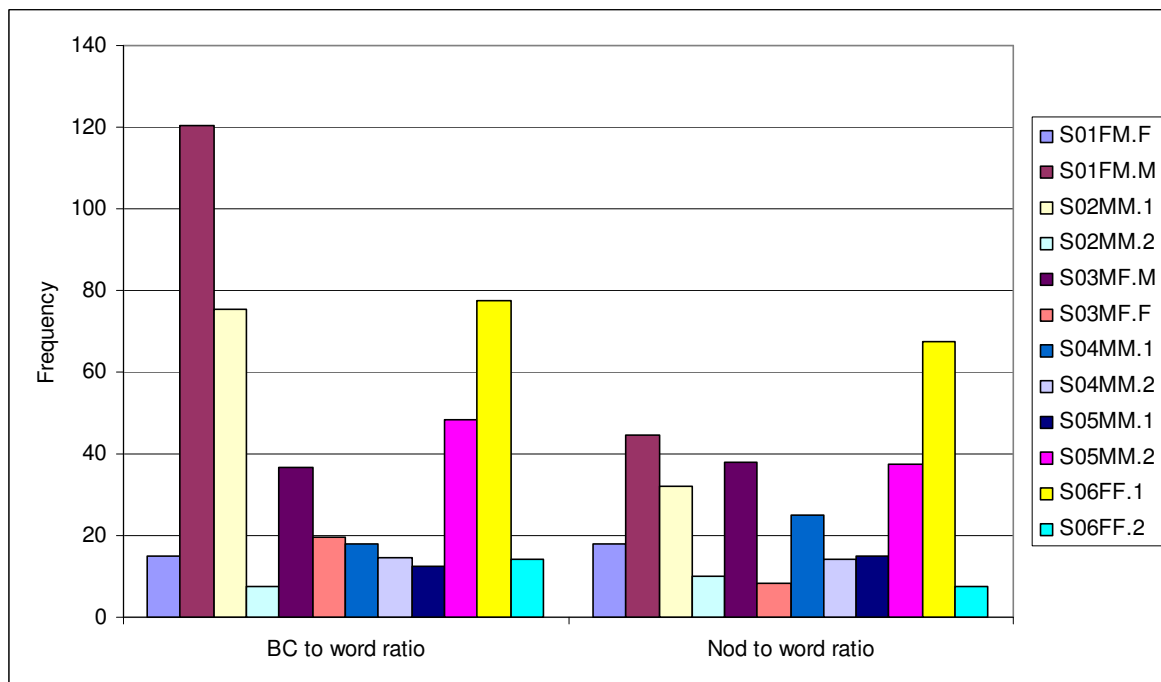
Figure 6.5: The ratio between word frequency and spoken / non-verbal backchannel usage across the five-hour corpus.

Furthermore, Appendix 6.14 presents the frequency with which, for example, different forms and types of spoken and non-verbal backchannels are used by each of the speaker's in the videos. Those numbers highlighted in pink represent the frequencies as a percentage proportion of use per speaker, from the total usage in each supervision, out of a total of 100%. So, for example, S01FM.F uses 3758 words from 8213 in S01FM, Appendix 6.14 thus shows that this speaker speaks 46% of the total word count whereas S01FM.M speaks 54%. It should be noted that the numbers in the light green

216

columns in Appendix 6.14 present the percentage of use of a particular backchannel as a proportion, a ratio, of the sum of spoken or non-verbal backchannels or specific forms or types of such behaviours as relevant, used by each individual speaker. As a result of this, the sum of the percentages given for the speakers in a particular supervision will not add up to necessarily 100% in these cases, as this figure is a percentage from the speaker's total rather than the video's total. For ease of reference, Appendix 6.15 presents a summary table of these conditions, marking, in each case, the speaker from each dyad who was seen to use the highest percentage of each condition listed in the table in Appendix 6.14.

If the results given for supervision S01FM are examined more closely, a notable disparity between the frequencies with which spoken backchannels are used by these two speakers becomes apparent. S01FM.F, the supervisor, uses a total of 250 spoken backchannels, whereas S01FM.M uses only 37. S01FM.F uses backchanneling nods more frequently than the other speaker, with 211 instances to only 100 by S01FM.M. Therefore, of the total 3754 words used by S01FM.F and the 4459 words used by S01FM.M, a ratio of 1:15 words to spoken backchannels and 1:18 words to non-verbal backchannels are used by S01FM.F, while only 1:121 and 1:45 are used by S01FM.M respectively, as charted in Figure 6.5. So, as a crude distinction, if the least frequent speaker is classified as the passive recipient, the listener, in this supervision, it is the listener who is seen to perform both spoken and non-verbal backchannels more frequently than the other participant. Furthermore, the 'speaker' uses more non-verbal backchannels than spoken forms, but for

the 'listener' this pattern is reversed. Although, interestingly, the proportion of spoken/ non-verbal backchannels used by these participants is fairly similar.

Similar results to these are seen in S02MM, as indicated in Figure 6.4. Here S02MM.2 uses a total of 469 spoken backchannels while S02MM.1 uses only 70, despite the fact S02MM.1 actually speaks much more than S02MM.2 throughout the video, with 5266 words to only 3502. This means that S02MM.1, the supervisor, uses spoken backchannels at a ratio of 1:75 words whereas S02MM.2 uses them at a ratio of 1:7 words, as demonstrated in Figure 6.5. Similarly, as a raw result, S02MM.1 is seen to use backchanneling nods less frequently than S02MM.2, with 165 to 350 instances. This amounts to ratios of 1:32 and 1:10 words to nods for S02MM.1 and S02MM.2 respectively. This means that S02MM.2 uses non-verbal backchannels at a rate of 2.3 times more often than spoken forms, whereas S02MM.2 uses spoken forms 1.34 times more frequently than nods.

In terms of the S03MF, a dramatic difference between the total numbers of words used by each speaker can be seen, with S03MF.M, the supervisor, speaking at more than twice the rate of S03MF.F, with 5834 to 2576 words. Nevertheless, there is less noticeable difference between the frequencies with which each speaker uses spoken backchannels than was seen in supervision S02MM. S03MF.M uses a total of 160 spoken backchannels, giving a backchannel to word ratio of 1:36, whereas S03MF.F uses 132, giving a ratio of 1:19. Conversely, S03MF.F uses backchanneling nods far more frequently than S03MF.M, with 311 instances to 154 instances by S03MF.F, i.e. nod to word ratios of 1:8 and 1:38 respectively.

There is only a small difference between the frequencies with which spoken backchannels are used by the participants in S04MM, so while 283 are used by S04MM.1, 250 are used by S04MM.2. This is despite the fact that S04MM.1 speaks more frequently than S04MM.2, with S04MM.1 using a total of 5066 words and S04MM.2 only 3610. This is matched by a spoken backchannels to word usage ratio which stands at 1:18 for S04MM.1 and 1:14 or S04MM.2. Similarly, there is a fairly consistent amount of backchanneling nods used by both the speakers in S04MM, with a total of 201 used by S04MM.1 and 252 by S04MM.2. This equates to a nod to word ratio of 1:25 for S04MM.1 and 1:14 for S04MM.2. In short, in this supervision video S04MM.2 uses more backchannels overall, but fewer spoken forms than S04MM.1, the supervisor, does.

S05MM has by far the largest word count of all the videos featured in this thesis, with 11338 words. The vast majority of these, i.e. 7032 words, are spoken by S05MM.2, the supervisee, rather than the supervisor, S05MM.1, who instead speaks a total of only 4306 words. Figures 6.4 and 6.5 indicate a dramatic difference between the frequencies of spoken backchannel use across the speakers, although this is in inverse relationship to the amount of words spoken by each participant. S05MM.1 performs more than double the amount of spoken backchannels than S05MM.2, with 342 instances to a mere 145, giving a backchannel to word ratio of 1:13 for S05MM.1 and 1:48 for S05MM.2. Similarly, S05MM.1 also uses more backchanneling nods than S05MM.2, with 286 to 187 nods, giving a ratio of 1:15 nods to words for S05MM.1 and 1:38 for S05MM.2.

As with S05MM and S03MF, there is a startling difference between the amounts of words spoken by each of the participants in S06FF, with the supervisor, S06FF.1, using 8154 words but S06FF.2 only using 2674 words. This is again inversely proportional to the amount of spoken backchannels used by the speakers, with S06FF.1 only using a total of 105 and S06FF.2 using 187; giving spoken backchannel to word ratios of 1:78 and 1:14 respectively, as shown in Figure 6.5. S06FF.2 also uses backchanneling nods more frequently and to a much greater extent than S06FF.1, almost 3 times, with 347 nods to 121 respectively.

In short, in all supervisions examined, the least frequent speaker uses fewer words to every spoken and non-verbal backchannel used, although, in terms of raw frequencies, in four of the six videos, including S01FM, S02MM, S05MM and S06MM, the participant who speaks less uses more spoken backchannels overall than the person who speaks most frequently throughout the supervision. The differences between the rates at which the participants use spoken backchannels in these particular videos are greater than the difference in the rate of use seen in the remaining videos, supervisions S03MF and S04MM. This suggests that the proportional 'constant' rate of backchannel and word use, as observed in Section 4.3.2.1 of the case study chapter, is not seen in this larger corpus.

In all instances, across the entire corpus, the least frequent speaker also uses a larger raw, net amount of spoken and non-verbal backchannels as a combined total. Despite this, there appears to be no clear patterns in the specific number of spoken backchannels and backchanneling nods used by

each speaker, as values for these are highly variable across each participant and each conversational episode.

Furthermore, there appears to be no clear relationship between the amount of words used and the professional status of the speaker. While in four videos, S02MM, S03MF, S04MM and S06FF, the supervisor speaks more than the supervisee, for the remaining two videos this relationship is reversed. Moreover, in three of the videos the supervisor uses more spoken backchannels than the supervisee, whereas, for the other three videos this is true of the supervisee.

6.3.1.2. Plotting the distribution of backchannel use

Further to these results, it is also relevant to note that backchanneling behaviours, in terms of spoken and non-verbal forms combined, proved to be used at a constant rate over time by all speakers in the corpus. Using the 'plot' facility in the Concord application of Wordsmith Tools (Scott, 1999), the distribution of a range of different characteristics of backchanneling behaviour over the course of a conversation can be graphically represented. These distributions are presented in the figures seen in Appendix 6.13, with each black mark representing the approximate juncture where a spoken and/or non-verbal backchannel occurs over time.

Plot 1 of Appendix 6.13 shows that backchannels are used fairly constantly used by each speaker in the corpus, as no marked differences in use is seen, for example, between the rate of use at the start, middle and end of each conversation. These findings prove to be consistent across spoken and non-verbal forms of backchanneling behaviour. Plots 2 and 3 of Appendix

5.13 show that although the frequencies with which these forms are used fluctuate across the speakers, their basic distribution over time is consistent for each participant.

However, again in reference to plot 1, it *is* necessary to mention that both speakers in S01FM and S02MM.1, as well as S03MF.M and S04MM.1 appear to use backchannels slightly more readily at the start of the conversation than at the end, whereas S02MM.2, S03MF.F, S04MM.2 and S05MM.1 use them slightly more at the end. This pattern is interesting as it suggests that in the majority of cases the supervisor uses slightly more backchannels at the start of the supervisions, while the supervisees use them at the end.

On the other hand, Plot 2 illustrates that there appears to be a clustering of backchannel use at certain intervals over time and across the two speakers in a dyad. Namely, there is a tendency for one speaker to use a single backchannel or a series of backchannels at a time when their partner does not. Therefore, it can naturally be assumed that at periods where a specific person is not backchanneling, they are most likely to be holding the floor at that given point in the conversation.

6.3.1.3. Summary

Overall, these initial findings can be summarised by the following points:

A. Spoken and non-verbal backchanneling behaviours are used at a near constant rate by all speakers across each conversational dyad. This finding supports premise 1.

B. There appears to be an inverse relationship between the number of words spoken by a participant and the rate at which s/he backchannels. So, the participant who speaks less backchannels more. This finding supports premise 2.

C. Consequently, those who speak less also have a greater nod-to-word ratio, so use a higher rate of nods to every word spoken.

D. However, those who speak less do not necessarily have a greater spoken backchannel to word ratio. Yet overall, these speakers are shown to use spoken and non-verbal backchannels collectively, at a higher rate, as a ratio to every word spoken, than the other speakers in the conversational dyad.

6.3.2. Spoken backchannels

6.3.2.1. Overview

In order to extend the discussions, it is now appropriate to focus specifically on the use of spoken backchannel forms. This line of enquiry essentially concentrates on 2 variables, the *type* and *function* of the backchannels and the relationship between these, according to the *frequency* with which they are used. Therefore, this addresses the question of whether a specific type or lexical form of backchannel adopts a specific function in the discourse more prevalently than other types or forms. Past studies exploring spoken backchanneling behaviour have generated the following findings. These will act as the premises fuelling investigations here:

3-   The simple backchanneling form *mmm* is most frequently used, based on results provided by Oreström, 1983; Gardner, 1997a, 1998 and O'Keeffe and Adolphs, 2008. The simple backchannels *yeah, okay* and *right* will also be fairly prevalent in talk, although they function in different ways to *mmm* (see O'Keeffe and Adolphs, 2008; Gardner, 1997b).

4-   The continuing (CON) function is most commonly adopted by spoken backchannels, as supported by Oreström, 1983 and O'Keeffe and Adolphs, 2008.

5-   Complex forms of backchannels commonly function in a more affective, relational way, than simple forms, such as *mmm* and *yeah.* These latter forms instead often function as continuer (CON) tokens, which are often more semantically empty; providing the 'most minimal' feedback (O'Keeffe and Adolphs, 2008).

The lexical form, functions and frequencies of spoken backchannel behaviours are documented in sections 1, 2 and 3 of Appendices 6.1 to 6.6 for supervisions S01FM to S06FF inclusive. Specific results, as discussed below, are also summarised in the relevant columns of the tables seen in Appendices 6.14 and 6.15.

6.3.2.2. Lexical structure

In terms of lexical structure, there is a wider range of backchannels of a 'complex' form used by S01FM.F in S01FM, than of any other form. This is not true of frequency, but rather in the mere range of spoken forms used. This

224

speaker uses 20 different complex forms of backchannels, and only 5 different 'simple' and 2 'double' forms; amounting to 74%, 18% and 8% of the total, respectively (refer to Chapter 2 for definitions). The widest range of spoken backchanneling forms used by S01FM.M is those of a simple type. Of the 14 different forms of spoken backchannels used by S01FM.M, 9 are classified as simple, 4 are complex and 1 is double (64%, 29% and 7%). Despite this disparity in the variety of forms, both speakers use the simple types of backchannels most frequently, with S01FM.F using 193 (77%) spoken backchannels as simple forms, 40 (16%) as complex forms and 17 (7%) as double forms and S01FM.M using 32 (86%), 4 (11%) and 1 (3%) respectively.

In S02MM, S02MM.1 uses 21 different lexical forms of backchannels in the supervision, the majority of which, 11 (52%), are simple form, while double backchannels have a frequency of 2 and complex forms have 8, representing 10% and 38% of the total respectively. In terms of frequency, again simple backchannels are used most often overall for S02MM.1, as seen on 56 different occasions across the video, so around 80% of instances. This is followed by complex forms and double forms, on 11 and 3 occasions, so representing 16% and 4% of the total number of spoken backchannels used by this speaker. Similar results are obtained for S02MM.2. Again, this speaker uses simple backchannel forms more readily than complex and double forms, comprising 74%, 13% and 13% of the total number of spoken backchannels used respectively, with cumulative frequencies of 348, 61 and 60. A wide range of different lexical forms is used overall by this speaker, with 8 different simple backchannels, 6 double forms and 30 different varieties of complex backchannels, thus representing 18%, 14% and 68% of the total respectively.

This stands in stark comparison to only 21 different lexical forms used by S02MM.1.

The results also indicate that in S03MF, simple forms are used on 98 of all instances by S03MF.M and 113 by S03MF.F, amounting to 61% of and 86% of the total number of spoken backchannels used by each speaker. Whereas, double forms represent 15% and <1% of the total for each speaker, with frequencies of 24 and 1, rates for complex forms are 24% and 14%, i.e. with frequencies of 38 and 18. The results revealed that there is a wide range of different lexical forms of backchannels used by each speaker; S03MF.M uses a total of 46 different forms, and S03MF.F uses 25. For both speakers, the majority of these are complex forms, with 35 different forms used by S03MF.M and 17 for S03MF.F, representing 76% and 68% of the total for each. These are followed by simple then double forms with frequencies of 8 and 3 for S03MF.M and 7 and 1 by S03MF.F, representing 17%, 7%, 28% and 4% of the total respectively.

In terms of supervision S04MM, there is a notable difference in the sheer variety of different forms used by each of the speakers, with S04MM.1 using 42, 10 of which (24%) are simple, 3, (7%), are double and 29 (69%) are complex forms. In comparison, S04MM.2 uses only 17 different forms, 8 (47%) are simple, 2 (22%) are double and 7 (41%), are complex. In terms of the frequency with which these forms are used in conversation, it is apparent that simple forms are again most often used by both S04MM.1 and S04MM.2, totalling 207 and 229 of the spoken backchannels used for each, representing 73% and 92% of the total used for each. These are followed by complex and double forms, with frequencies of 38 and 7 for S04MM.1 and 11 and 10 for

S04MM.2, amounting to 25%, 2%, 4% and 4% respectively of the totals for each of these speakers.

By contrast, there are a fairly even proportion of simple, double and complex spoken forms used across the speakers in S05MM. Of the 27 different forms used by S05MM.1, 8 are simple, 3 double and 16 are complex. Whereas for S05MM.2, of the 13 different forms used, 5 are simple, 1 is double and 7 are complex. Despite the fact there is less variety in this supervision, both speakers consistently use simple forms at a higher rate than the other forms. This amounts to 93% of the total both for S05MM.1 and S05MM.2, 317 and 135 from 342 and 145 respectively. These are followed by complex, then double, forms, amounting to 6% and 1% respectively for S05MM.1, and frequencies of 21 and 4, and 6% and 1% for S05MM.2, so with frequencies of 8 and 2.

Despite using more spoken backchannels, S06FF.2 uses a smaller range of lexical forms of these than S06FF.1, a mere 16. 10 of these are classified as simple forms, 0 double and 6 complex, with these representing 63%, 0% and 17% of the total. In contrast, S06FF.1 uses a total of 21 different structural forms, 6 of which are simple, 1 double and 14 complex ones (29%, 4% and 67% of the total, respectively). Again, in terms of frequency the rate at which simple forms are used is higher than the other forms of spoken backchannels, representing 86% (89) and 96% (180) of the total use respectively for S06FF.1 and S06FF.2, while complex forms represent 14% (15) and 4% (7) of the total and double <1% (1) and 0% (i.e. with a frequency of 0).

Across the complete corpus sample, it is evident that simple form backchannels are most frequently used overall. This is followed by complex forms, then double forms, although the difference between the uses of these is less marked then the difference between these forms and simple spoken backchannels. However, a wider variety of complex lexical forms of spoken backchannels are used than other forms, as the most common simple forms used from one speaker to the next appears more standardised and consistent.

6.3.2.3. Lexical form

The combined frequencies of the top ten most often used spoken backchannels seen in the corpus, in terms of lexical form, are presented in Figure 6.6. Complete lists of spoken backchannel forms, and corresponding frequencies across each supervision and the complete dataset, can be found in Appendices 6.7 and 6.8.

Rank	Lexical Form	Freq.	Rank	Lexical Form	Freq.
1	Mmm	793	6	Okay	81
2	Yeah	672	7	Mmm mmm	59
3	Yes	167	8	Yeah yeah	57
4	Right	116	9	Sure	34
5	Mhm	103	10	Uh hm	24

Figure 6.6: The 10 most frequent spoken backchannel forms in the corpus.

Appendix 6.1 indicates that the most commonly used lexical forms for both speakers in S01FM are the responses *yeah* and *mmm*. These are used collectively by S01FM.F a total of 182 times (73%) from the speakers' total

number of 250 spoken backchannels, and 26 times from a total of 37 by S01FM.M (65%). Similarly, the most common lexical form of spoken backchannels used by S02MM.2 is *yeah*, which is used a total of 16 times whereas *mmm* is used on only 1 occasion. This pattern is reversed for S02MM.1, who uses *mmm* 243 times and *yeah* only 56 times. *Yeah* is also the most frequently used form by both speakers in S03MF, 53 times by S03MF.M and 75 times by S03MF.F. *Okay*, *right* and finally *mmm* all feature in the top ten most frequently used backchannels by these speakers, so unlike S02MM there is quite a high correlation between spoken backchannel used by each of the speakers in this particular supervision.

As with S02MM.1, S04MM.1 also uses the simple form *mmm* most frequently, with 104 occurrences; a proportion of 37% of all spoken backchannels used by this speaker. Although, this is closely followed by *yeah,* with 83 uses, (29% of the total). Whereas S04MM.2 again most commonly uses *yeah,* with 149 occurrences (60% of the total), using *mmm* only 48 times (19% of the total). Despite differences in the raw frequencies of these forms, both *mmm* and *yeah* thus remain the most commonly used across the speakers in S04MM.

In contrast, the third most common spoken backchannel used by S04MM.2, the lexeme *right,* which occurs 16 times, 6% of all instances, is only used on 2 occasions by S04MM.1, so at a rate of less than 1%. Conversely, the lexeme *yes* is used on only 2 occasions by S04MM.2, again at a rate of <1%, whereas it is used 25 times by S04MM.1, accounting for about 9% of the total number of spoken backchannels uttered by this speaker.

Again, the same two forms *mmm* and *yeah* are listed within the top three most commonly used for both speakers in S05MM. *Mmm* and *yeah* have frequencies of 179 and 27 (52% and 8% of the total, respectively), for S05MM.1 and 89 and 41 for S05MM.2 (61% and 28% of the total). Again, it is interesting to note that the second most common form for S05MM.1 is *yes*, accounting for 27% (92) of all spoken backchannels used by this speaker, whereas for S05MM.2 *yes* is <1%, with a frequency of 1.

In final supervision video S06FF, the results reveal that the lexical forms *mhm* (rather than *mmm*) and *yeah* are by far the most commonly used by S06FF.2, amounting to 140 (75% of the total for this speaker). In contrast, for speaker S06FF.1 the forms *right* and *yes* are the most common, as used in 55% of all occasions, a frequency of 58. However, *yeah* follows closely for S06FF.2, with a frequency of 21, so representing 20% of the total for this speaker.

Overall then, Figure 6.6 and the appendices listed above, support the proposition given in premise 3, insofar as the simple form backchannel *yeah* appears within the top three most frequent spoken backchanneling forms for all of the speakers. This result differs to that seen in the case study, where *mmm* was the most common lexical form of spoken backchannel used, see 4.3.2.2 for further details. In addition, in the corpus, *mmm* appears in the top five most frequent spoken backchanneling forms used by all speakers apart from S02MM.1, who only uses this response on one occasion.

The spoken backchannel *okay* also appears within the top ten most frequently used backchanneling tokens for all but one of the speakers, S06FF.1, whereas *right* appear in the top ten backchannels for 8 of the 12

speakers, while *yeah yeah* appears in the top ten for 7 speakers. Of all the speakers, S03MF.F and S01FM.M use the least amount of these most frequent forms, with only 4 appearing in their respective top ten most frequent backchannel lists. Whereas S04MM.2 and S05MM.2 use the greatest amount, each with 7 of these appearing within the top ten most frequently used by these speakers.

6.3.2.4. Function

Figure 6.7 provides a breakdown of the functions that spoken backchannels most commonly adopt in this dataset. This allows the investigation of premise 4 to be undertaken in greater detail (again refer to Appendices 6.7 and 6.8 for combined tables of these results).

Speaker	Discourse Function				Speaker Total	Grand Total
	CON	CNV	ER	IR		
S01FM.F	163	68	9	10	250	287
S01FM.M	14	18	2	3	37	
S02MM.1	13	33	10	14	70	539
S02MM.2	345	93	24	7	469	
S03MF.M	51	50	17	42	160	292
S03MF.F	52	61	8	11	132	
S04MM.1	157	89	24	13	283	533
S04MM.2	132	102	4	12	250	
S05MM.1	197	120	16	9	342	487
S05MM.2	97	38	6	4	145	
S06FF.1	17	56	7	25	105	292
S06FF.2	97	61	4	25	187	
	1335	789	131	175		2430

Figure 6.7: The functions and frequencies of spoken backchannels in the corpus.

In terms of supervision S01FM, this figure reveals that while S01FM.F uses backchannels functioning as CON tokens most frequently, amounting to 65% of all instances, this stands at only 38% for S01FM.M; refer to the above figure for specific details of frequency. S01FM.M instead uses CNV tokens most frequently, as seen in 49% of all instances, while this proportion stands at only 27% for S01FM.F. Both speakers use the IR and ER tokens the least with only a small difference between the frequencies with which each of these functions are used. IR and ER tokens are both used in *circa* 4% of instances by S01FM.F, whereas for S01FM.M this stands at 5% and 8% respectively.

The most common discourse function of backchannels used by S02MM.1 is CNV tokens (47% of all instances), while forms adopting a more CON function are relatively infrequent (19% of the total). Since 20% of these backchannels function as IR tokens for this speaker, the CON function exists as only the third most frequently used by him. This result is strikingly different to that seen for S02MM.2 who uses spoken backchannels with a CON function on 74% of all instances, and CNV on only 20%, where ER and IR tokens comprise only 5% and 1% of the total.

The most frequently used spoken backchannels in S03MF are again CON and CNV tokens. This is true for both speakers, with S03MF.M using them on 32% and 31% of all occasions, and S03MF.F 39% and 46% respectively. There is a disparity between the uses of IR tokens across the speakers, with S03MF.M using them in 26% of all instances, whereas for S03MF.F this stands at only 7%. This result coincides with what was seen across videos S01FM and S02MM, both with speaker 1, the supervisor, using a larger amount of IR tokens than the other speaker does. However, the percentage of

use for S01FM.F is quite low owing to the sheer amount of backchannels used by this speaker.

As with the previous supervisions, the most frequent discourse functions adopted by the spoken backchannels used in S04MM are CON and CNV tokens. Of the total 533 spoken backchannels used in the video, 54% are CON and 36% CNV tokens, with a similar proportion of these used by the individual speakers. Similarly, in terms of S05MM, while S05MM.2 uses CON tokens more than S05MM.1, with 67% to 58% of the total backchannels usage for each speaker, S05MM.1 uses CNV tokens more than S05MM.2, with 35% to 26%. There is a negligible difference between the frequency with which ER and IR tokens are used by each speaker.

Finally, results for S06FF reveal that spoken backchannels functioning as IR tokens are actually used on 24% of all occasions by S06FF.1, making this function the second most frequent adopted by this speaker. This is followed by backchannels functioning as CNV tokens, which account for 53% of all spoken backchannels used by this speaker. For S06FF.2, the use of IR tokens stands at only 13%; the third most commonly used function for this speaker. The most common function used by S06FF.2 is, again, the CON function, accounting for 52% of the total for this speaker, followed by CNV backchannels, at 33%. Collectively, the speakers use backchannels as ER tokens least frequently, amounting to 7% of the total used by S06FF.1 and 2% of the total for S06FF.2.

In short, these results demonstrate that the most frequently used response tokens across all of the supervisions are CON backchannels, followed by CNV, IR and ER tokens. This pattern parallels the results seen in the case

study analysis, and supports premise 4 of the thesis (refer to 4.3.2.3 for further details). While this pattern is true of eight individual speakers (S01FM.F, S02MM.2, S03MF.M, S04MM.1, S04MM.2, S05MM.1, S05MM.2 and S06FF.2), S02MM.1 and S06FF.1 use CNV tokens most frequently, followed by IR, CNV and ER tokens. The remaining speakers, S01FM.M and S03MF.F use spoken backchannels functioning as CNV most often, followed by CON, IR and ER tokens. None of the speakers use spoken backchannels functioning as ER or IR tokens most frequently.

Across the six supervision videos there is a fairly even proportion of CON and CNV tokens used by each individual, when considered as a percentage of the total amount of spoken backchannels used by each of them (refer to Appendices 6.7 and 6.8 and Figure 6.4). There is a more marked difference between the raw frequencies with which these functions are used across the speakers. It is interesting to note that the least active speakers, in terms of word-count from each of the supervisions (see Section 6.3.1, above), use spoken backchannels functioning as CNV tokens more frequently than the other participant, as a net amount. This is true in 100% of the cases.

In contrast, the most active speaker uses a higher proportion of their spoken backchannels as CNV tokens than the other speaker does. This again is seen in 100% of the supervisions. In 83% of these cases the least active speaker uses a higher net amount of CON and ER backchanneling tokens than the other speaker, whereas IR tokens are more evenly distributed across the least and most frequent speakers.

In five out of the six supervisions, with S02MM as the exception, the supervisor uses more spoken backchannels functioning as ER tokens than

the supervisee, while in all six of the videos the supervisor uses either the same amount, as is the case for S06FF, or more spoken backchannels functioning as IR tokens, than the supervisee. Both of these results are in terms of the total net amount of ER and IR tokens used in each of the supervisions.

In respect of the proportion with which each speaker uses these tokens, the results indicate that on five out of six occasions the supervisors use a higher proportion of spoken backchannels as IR tokens than the supervisees do. This is with the exception of S01FM. The difference in the use of spoken backchannels as ER tokens is not as marked between supervisors and supervisees, as in three of the videos (S01FM, S03MF and S06FF), the supervisor uses a higher proportion of ER tokens than the supervisee, whereas for the remaining three this pattern is reversed.

6.3.2.5. The relationship between forms and functions

A summary of the relationships between spoken backchannel form and function, based on the most common forms from Figure 6.6, is charted in Figure 6.8.

Unsurprisingly, the most frequently used discourse function adopted by these spoken backchannels is the CON function, followed by CNV. This is with CON being used around twice the number of times compared to CNV tokens representing 61 % and 33% of the grand total of 2106, i.e. frequencies of 1282 and 686 respectively. IR tokens are the third most commonly used function with these spoken forms, with 133 occurrences, whereas for ER there are only 5 occurrences, representing 6% and <1% of the grand total.

Lexical Form	Discourse Function				Lexical Form	Discourse Function			
	CON	CNV	ER	IR		CON	CNV	ER	IR
Mmm	793	0	0	0	Mmm mmm	59	0	0	0
Yeah	293	378	0	1					
Yes	1	165	0	1	Yeah yeah	8	49	0	0
Right	1	40	2	73					
Mhm	101	2	0	0	Sure	1	27	3	3
Okay	1	25	0	55	Uh-hm	24	0	0	0

Figure 6.8: Mapping the most common functions of the most frequent spoken backchannel forms in the corpus.

In this corpus, the most frequently used CON tokens are the backchanneling forms *mmm*, *mhm*, *mmm mmm* and *uh hm* respectively accounting for 60%, 8%, 4% and 2% of *all* 1335 CON tokens used. In contrast, *yeah, yes, yeah yeah* and *sure* most frequently function as CNVs, representing rates of 48%, 21%, 6% and 3% of the total. *Right* and *okay* are most commonly used in the role of IR tokens, amounting to 42% and 31% of all IR tokens seen. None of the top ten spoken backchannel forms in this corpus are most frequency used to function as ER tokens.

In terms of individual speakers, in S01FM the backchannel *yeah*, used as a CNV, is the most frequently used for S01FM.M. This is true of 14 instances, 38% of all spoken backchannels used by this speaker. Whereas, for S01FM.F the most common form is *mmm*, used as a CON, as seen on 103 occasions, (41% of the total). The second most common backchannel for this speaker is again *yeah* functioning as a CNV, as used on 45 occasions, (18% of the total). *Okay* most frequently functions as an ER by both speakers in this supervision, although the frequencies for this are relatively low, with rates of 3 and 2, (1%

and 2% of the total number of spoken backchannels used by both speakers). No single backchannel form/ expression is used more than once as an IR token by either of the speakers in this supervision.

By contrast a large quantity of the IR tokens are used by S02MM.1. The lexeme *right* is used as an IR token on 6 occasions by this speaker, and 10 times in total across all discourse functions,(14% of the total). However, this is only used twice by S02MM.2, once as an IR and once as a CNV, thus representing <1% of the total. A similar situation exists for the lexeme *okay* which is only used on one occasion by S02MM.2, as an IR, yet 8 times by S02MM.1, making this the fourth most frequent spoken token used by this participant, accounting for 11% of all backchanneling tokens spoken by S02MM.1 and <1% for S02MM.2. Of the ER tokens used, the spoken backchannels *right*, *oh right* and *good* are the most commonly adopted by S02MM.1, each at a frequency of 2 (3% of all spoken backchannels used by this speaker). *Definitely* is used 4 times by S02MM.2, representing <1% of the grand total, although this amounts to 17% of all ER tokens used by this participant.

For S02MM.1 and both speakers in S03MF, the lexical item *yeah* is the most commonly used CON token, whereas for S02MM.2 it is the form *mmm*. The backchannel most often used as a CNV is the simple form *yeah,* that is, 45% of the total number of backchannels functioning as CNV tokens used by this speaker, with a frequency of 42. Similar to S02MM.1 (frequency of 23), the lexeme *right* is the spoken form that most often functions as a IR backchannel for S03MF.M, representing 56% of the total number of IR tokens used by this speaker. This is followed by *okay,* with a frequency of 6 (14% of

the total). This pattern is reversed for S03MF.F as of all IR tokens used (45%), 5 are the response *okay* and 3, (27%), are the lexeme *right*. Again, there are no specific lexical forms that are frequently used as ER tokens in this supervision; instead all forms adopting this function have a frequency of only 1 instance. This is also true for S04MM.1 and S04MM.2.

Yeah is again the most common backchannel functioning as a CON token for S04MM.2 and *mmm* is the second most frequently used in this way, with frequencies of 76 and 48, i.e. 59% and 36% of the total number of CON tokens used by this speaker. *Mmm* is the most common backchannel used as a CON token by S04MM.1 and, conversely, *yeah* is the second most commonly used. These have frequencies of 104 and 42 respectively, thus represent 66% and 27% of the total. *Yeah* functions as a CNV more times than any other backchannel for both of these speakers, comprising 46% and 72% of all CNV tokens used for S04MM.1 and S04MM.2 respectively, with frequencies of 41 and 73. *Okay* is often used as an IR across this supervision, although this is more frequently the case with S04MM.1 than S04MM.2, who instead uses the lexical item *right* as an IR a total of 8 times, so 67% of all occasions where a spoken backchannel functions as an IR token. Whereas, *okay* is used in this way on 4 occasions, 23% of all instances of IR use. This stands at a rate of 54%, i.e. a frequency of 7, by S04MM.1.

For both speakers in S05MM, as well as S06FF.2, it is *mmm* that is most often used as a backchannel functioning as a CON token, with frequencies of 179, 89 and 6, thus representing 91%, 92% and 35% of the total number of CON tokens used by S05MM.1, S05MM.2 and S06FF.2 respectively. *Mhm* is used most often in this way for S06FF.1, as seen in 83 occasions, 86% from

the total of 97 CON tokens used by this speaker. Furthermore, S05MM.2 and S06MM.2, the supervisees, both use *yeah* as a CNV token more frequently than any other spoken backchannel, comprising 87% and 84% of all CNV backchannels used by these speakers, with frequencies of 33 and 51. This is in contrast to both S05MM.1 and S06FF.1, the supervisors, who use the more standard lexical item, *yes*, more often in this way than they do *yeah.* This is true for 92 and 25 instances, comprising 77% and 45% of the total CNV tokens used by each. *That's right* and *yeah absolutely* are the spoken forms, used by S05MM.1 and S05MM.2 which most commonly function as ER tokens, with frequencies of 3 and 3, comprising 19% and 50% of all ER tokens used by them. The ER token use in S06FF is possibly too minimal for comment.

Okay and *right* often function as IR tokens across all speakers in S05MM and S06FF. *Right* is used 22 times as an IR token by S06FF.1, 88% of all IR tokens, whereas *right* is only used 6 times by S06FF.1 but *okay* is used 18 times, so at 24% and 72% of the total number of IR tokens used by her. Finally, in S05MM.1 uses *okay* as a IR on 56% of all uses of IR tokens, i.e. 5 times, whereas S05MM.2 uses *right* most often as an IR token, but at a fairly minimal rate, i.e. 2 occasions, nevertheless, this amounts to 50% of all IR tokens used by this speaker.

6.3.2.6. Summary

Building on observations in Section 6.3.1.2, the findings generated from the investigations in 6.3.2 are summarised overleaf:

E. Simple form backchannels, comprising of a single lexical item are far more commonly used than double or complex forms. These are the 'most minimal' forms of spoken backchannels.

F. Conversely, there is more variety in the lexical structure of complex forms of backchannels, in other words there is a larger range of complex than simple forms.

G. Simple form spoken backchannels are most often used as CON and CNV tokens, except for the *right* and *okay*. This result supports premise 4. It is important to note, however, that some of these forms were more strongly associated with one common function than others*.*

H. *Mmm* and *yeah,* and derivations such as repetitions in double forms, and the non-standard form of *yeah*, *yes*, are the most commonly used spoken backchannels in dyadic conversation, with *mmm* acting as a CON and *yeah* as a CNV in the majority of instances. This result supports premise 3.

I. CON and CNV functions are most commonly used across the entire corpus. ER tokens are the least common.

J. Complex and double forms are used to fulfil the function of ER and IR tokens, i.e. they adopt more affective roles, more often than simple forms. This result supports the implication contained in premise 5.

6.3.3. Non-verbal backchannels

6.3.3.1. Overview

In order to explore the patterns of backchanneling nod usage across the corpus, the following statements, which were devised with reference to the case study findings in 4.3.3, will be investigated:

6-   Backchanneling head nods are used at the same rate or more *frequently* than spoken backchannels since they are even more minimal and non-evasive than spoken forms, imposing even less of a challenge to the floor.

7-   The most common *types* of head nods used in discourse are of a short duration, i.e. types **A** and **C** or less intense, multiple, type **B** nods. Types **D** and **E** are less frequently used.

Details of non-verbal backchannel use can be found in sections 1, 2, 4 and 5 of Appendices 6.1 to 6.6 for supervisions S01FM to S06FF inclusive.

6.3.3.2. Nod type

Figure 6.9 charts the frequency with which the 5 different backchanneling nods types are used by each of the speakers featured in the five-hour dataset. This includes nods that are used with and without concurrent spoken backchannels, see Section 6.3.4 for more specific explorations of behaviour according to these categories.

This figure indicates that S01FM.F uses type **A** nods more frequently than any other type, as seen in 31% of all instances, refer to Figure 6.9 for details

of specific frequencies. This is followed by types **B**, **C**, **D** and **E**. Whereas, S01FM.M uses types **A** and **B** at a similar rate, each amounting to 61% of the total number of backchanneling nods used. These are followed by types **C**, **D** and **E**.

Speaker	Nod Type					Speaker Total	Grand Total
	A	**B**	**C**	**D**	**E**		
S01FM.F	65	66	56	16	8	211	311
S01FM.M	61	28	10	1	0	100	
S02MM.1	94	18	41	7	5	165	515
S02MM.2	126	129	42	11	42	350	
S03MF.M	66	40	32	8	8	154	465
S03MF.F	203	30	74	1	3	311	
S04MM.1	82	40	37	33	9	201	453
S04MM.2	144	75	26	3	4	252	
S05MM.1	176	72	31	2	5	286	473
S05MM.2	137	11	37	1	1	187	
S06FF.1	44	22	32	13	10	121	468
S06FF.2	58	214	27	24	24	347	
	1256	**745**	**445**	**120**	**119**		**2684**

Figure 6.9: The types and frequencies of non-verbal backchannels in the corpus.

Type **A** nods are also most frequently used by S02MM.1, in 57% of all occasions. While type **B** nods are most frequently used by S02MM.2; amounting to only 37% of the total, although this is closely followed by type **A** nods, with 36% of the total amount of nods used. Type **D** nods are used infrequently by S02MM.2, at only 3% of the total, while types **C** and **E** are each used on 13% of all instances. Type **B** nods are used less often by S02MM.1 than S02MM.2 and type **C** nods are used more than twice the amount of times than type **B** by this speaker, comprising 25% of all nods,

compared to 11% for type **C**. Types **D** and **E** nods are used the least frequently by S02MM.2.

Type **A** nods are most frequently used in S03MF, representing 43% of all nods used by S03MF.M and 65% of those used by S03MF.F. These are followed by nod types **B** then **C** for S03MF.M, and types **C** and **B** for S03MF.F. The least frequently used nods for both speakers are of types **D** and **E**, each amounting to 5% and 5% use for S03MF.M and <1% and 1% for S03MF.F respectively. So, similar patterns of backchanneling head nod behaviour are seen across both speakers.

S04MM.2 is shown to use type **A** rather than type **B** nods in the highest proportion of cases, 88% of the total instances, 56% (61) for type **A** and 32%, (34) for type **B**. While this is also true for S04MM.1, the combined total percentage of use for these types is much lower, at only 61%, with 32% for type **A** and 29% for type **B**. There is only a slight difference in the use of type **C** nods across these speakers, 18% for S04MM.1 and 10% for S04MM.2, yet a greater difference in the use of types **D** and **E** nods, representing 16% and 4% of the total for S04MM.1, 1% and 2% for S04MM.2.

The patterns of nod usage in S05MM and S06FF are consistent with what has been seen thus far, with type **A** nods the most frequent for the majority of speakers featured in these videos, used on 62%, 73% and 36% of all instances by S05MM.1, S06FF.1 and S05MM.2 respectively. The only exception to this is S06FF.2 who, in contrast, uses type **B** significantly more frequently than other types, representing 62% of the total number of nods for this speaker, while type **A** nods are only used on 17% of all instances.

Both speakers in S05MM speakers use type **D** and **E** nods the least frequently, although S05MM.1 uses type **B** nods significantly more often than type **C** nods, each at 25% and 11% of the total, while S05MM.2 uses significantly more type **C** nods than type **B** nods, at rate of 20% and 6%. In short S05MM.1 most frequently performs backchanneling nods of a low intensity in this conversation, whereas S05MM.2 uses short duration nods.

In S06FF, type **C, D** and **E** nods, in this order, are the least frequently used by S06FF.2, seen on only 8%, 7% and 7% of all occasions, while types **B**, **D** and **E** are the least frequently used nods used by S06FF.1, as seen on 18%, 11% and 8% of all instances.

Overall, S06FF.2, S02MM.2 and S03MF.F have been shown to use the most backchanneling nods in the corpus, however, since these speakers are featured in videos that are, on average, the longest length (refer back to Figure 6.4) such a result is not particularly significant. What is interesting to note is that in 100% of the supervisions, the least frequent speaker, i.e. the most frequent 'nodder', uses more type **A** nods than the other speaker. In addition, in 5 from 6 of the videos (83%), the least frequent speakers uses, on average, more type **B** nods than the most frequent speakers, except for S03MF where this trend is reversed. In four of the six supervisions (67%) the infrequent speakers, those adopting the role of the passive listener, also used more type **D** and **E** nods than the other participant, except for supervisions S03MF and S04MM.

In all of the videos the supervisors use a higher percentage of their non-verbal backchannels as type **D** nods, than the supervisees do (although this percentage is a meagre 1% for S05MM). This is also true for type **E** nods in

all of the supervisions examined, except for S02MM. For nod types **A, B** and **C** there is a less marked difference between the amount used by the supervisors and supervisees

6.3.3.3. Summary

The backchanneling head nods featured in this corpus are shown to adopt the following patterns of behaviour:

K. No clear relationship exists between, purely, the number of head nods performed and the number of spoken backchannels used by a speaker. This refutes premise 6. Although in some cases backchanneling nods are used at the same rate/more frequently than spoken forms, in 50% of cases, nods were less frequent.

L. In terms of the *individual nod types* used across the corpus, the results indicate that types **A, B, C, D** and **E** are the most frequently used, and in this order, although this sequence differs across individual speakers. In other words, the less intense nods, both of a long and short duration, were used more often than more intense and variable nods. There was also a tendency for nods of a shorter duration to be used more often than those of a longer duration. Type **A** nods are the most common overall. This finding supports premise 7.

6.3.4. Combining spoken and non-verbal behaviours

6.3.4.1. Overview

The next phase of enquiry examines the closeness of the relationship between spoken and non-verbal backchannel usage in more detail. For this, it is appropriate to test to what extent the following statements are true. These are again based on findings derived from the case study, see Chapter 4 for further details:

8- Nods are used more frequently with concurrent spoken backchannels than alone. Similarly, spoken backchannels are used more frequently with concurrent nods than alone.

9- Spoken backchannels that are used as IR and ER tokens are more likely to co-occur with complex forms of backchanneling nods that vary with intensity, i.e. types **B**, **D** and **E,** whereas backchannels that exist on the opposite end of the 'functional cline' will co-occur with shorter, more simple, type **A** and **C** nods.

Simple spoken forms that co-occur with nods of a low intensity and/or short duration are generally seen to be used in the same way as when a simple lexical form is used on its own. So, these are often performed, for example, at a TRP (see Section 2.3.1.3 of Chapter 2), providing minimal feedback between the speech of the speaker without interrupting or dramatically overlapping their speech. In such cases, the nod starts at the same time as the concurrent spoken form and ceases before or at the same

time, giving a one-to-one relationship between the spoken and non-verbal backchanneling behaviours.

However, there are many instances where this one-to-one relationship does not exist. In other words, on occasions where spoken backchannels co-occur with backchanneling nods, the timing of the backchannels used is not necessarily consistent. Since the length of a backchanneling nod is generally more variable than a spoken backchannel, as the length of the latter is dependent on the lexical form (even complex forms tend to be only up to a maximum of 6 or 7 words in length), on occasion a listener may start a backchanneling nod prior to uttering a concurrent spoken backchannel. Similarly, it is possible that the nod may continue for a time after the spoken form has been delivered.

This can be described as *nodding across turn boundaries*. This phenomenon is logically hypothesised to be particularly characteristic of nods with a longer duration, such as types **B, D** and **E,** and is explored in more detail in 6.3.5.

An example of this can be seen in the transcript excerpt taken from S02MM, presented in Figure 6.10. In this instance, although two different spoken backchannel forms, *yeah* and *yeah*, are used in successive turns by <$M2>, they are used at the same time as a single backchanneling type **B** nod which stretches over all turn boundaries, rather than two different, individual nods.

```
<$1> +which are usually just signifiers of bourgeois wealth+
<$2> Yeah,
<$1> +they're not usually it's just objects it's things+
<$2> Yeah,
<$1> +it's stuff or it's inns which are kind of transitional minimal
```

Figure 6.10: Nodding across multiple spoken backchannels and turn boundaries.

This phenomenon is seen in many of the videos included in this five-hour corpus sample. Sections 6 of Appendices 6.1 – 6.6 provide not only details of 'backchannels across multiple turns', as identified above, but:

- The raw number of spoken backchannels that are used with a nod at some point along the nod's duration. So, for the example above, it would be documented that in 2 cases backchanneling CNV tokens are used with a nod.

- The raw number of backchanneling nods that are used with at least one spoken backchannel. So, for the example above, 1 nod would be documented.

Before proceeding to explore the phenomenon of 'nodding across turn boundaries' in more detail, the following section examines the basic relationship between the co-occurrence of spoken and non-verbal forms, providing distinct totals of spoken and non-verbal backchannel co-occurrence and non-verbal and spoken backchannel co-occurrence. Following this,

Section 6.3.5 provides a more detailed investigation of nods that co-occur with more than one spoken backchannel in the corpus.

6.3.4.2. General results

Figure 6.11 charts the number of nods that co-occur with spoken backchannel forms, and, conversely, the number of spoken forms that co-occur with backchanneling nods. It is necessary to note that for the purpose of this table, a nod which is used across a number of turns and/or spoken backchannels is counted as a single nod. Therefore, the results seen in the 'nod with spoken backchannels' column do not match directly to the 'spoken backchannels to nod' column, given that a single nod can be used with more than one spoken form, across turns.

Speaker	Spoken Forms		Total	Nods		Total
	+ Nods	No Nod		No Spoken	+ Spoken	
S01FM.F	151	99	250	144	67	211
S01FM.M	20	17	37	20	80	100
S02MM.1	52	18	70	50	115	165
S02MM.2	379	90	469	262	88	350
S03MF.M	97	63	160	83	71	154
S03MF.F	106	26	132	106	205	311
S04MM.1	180	103	283	139	62	201
S04MM.2	160	90	250	144	108	252
S05MM.1	260	82	342	233	53	286
S05MM.2	110	35	145	110	77	187
S06FF.1	89	16	105	78	43	121
S06FF.2	133	54	187	129	218	347
	1737	**693**		**1498**	**1187**	

Figure 6.11: Frequencies of spoken and non-verbal backchannel co-occurrence across the corpus.

This figure indicates that in all videos spoken backchannels are used more frequently, at nearly twice the rate, with concurrent backchanneling head nods, with a frequency of 1737, than in isolation (see the '- nods' column). This accords with preliminary results seen in Section 4.3.2.1 of the case study chapter. The relationship between nods and concurrent spoken forms appears to be less consistent. Although 1737 spoken backchannels co-occur with a nod, the figure suggests that total of 1498 different nods were used with 1737 different spoken backchannels, while a total of 1187 different nods were used alone, without a spoken counterpart. It should also be noted that in 100% of all instances examined, the participant who speaks the least in each supervision video, also performs more concurrent non-verbal and spoken backchannels than their more 'vocal' counterpart.

Specifically in S01FM, S01FM.F uses non-verbal backchannels with spoken forms at a more frequent rate than she uses nods alone, with proportions of 68% to 32%. Whereas S01FM.M uses nods more frequently in isolation, as seen on 20% of all instances, compared with 80% for nods with concurrent spoken forms. Refer to the '+ spoken' and '- spoken columns' in Figure 6.11 for numerical frequencies of these states.

Similarly, in supervisions S02MM, S03MF and S06FF, while S02MM.2, S03MF.M and S06FF.1 use a greater proportion of their nods with spoken backchannels than without, the remainder of the speakers use more nods in isolation than with spoken forms (all do so with a proportion of around >2 times more than with concurrent spoken forms). S02MM.1 uses 70% of all non-verbal backchannels with concurrent spoken backchannels, S03MF.F uses 66%, and S06FF.2 63%. In the case of S03MF.M, nods with spoken

forms are used in 54% of all instances, whereas those without are used in 46%, so a mere 4% difference in the proportional rate of use. S02MM.2 uses backchanneling nods with concurrent spoken forms in 75% of all instances and S06FF.1 uses nods with spoken backchannels in 54% of the total.

In both S04MM and S05MM, nods are used more often with spoken forms than in isolation. For S04MM.1, 69% of the backchanneling are nods co-occurring with spoken backchannels, but this is a less frequent 57% for S04MM.2. Similarly, S05MM.2 uses 59% of all backchanneling nods with concurrent spoken forms, yet this proportion is 81% for S05MM.1.

As an extension to this line of investigation, it should be mentioned that the positions at which spoken and non-verbal backchannels co-occur or not across the stretch of the discourse have been plotted in plots 4, 10 and 11 of Appendix 6.13. Plots 12-16 provide a breakdown of the intervals at which spoken backchannels are used concurrently with each specific type of head nod across each conversation. Again these results illustrate that there is no marked difference in backchannel use, according to these three states, i.e. nods without spoken counterparts; spoken forms without nods; concurrent spoken and non-verbal backchannels respectively across each speaker/ supervision. In other words, no marked differences in the use of spoken and/or non-verbal backchannels appear over the course of a conversation. However, natural fluctuations in such behaviours do occur from person to person, and there is no consistency in frequencies.

6.3.4.3. Backchanneling nods with spoken backchannels

Figure 6.12 provides a breakdown of the frequency with which individual backchanneling nod *types* are used with concurrent spoken backchannels. Again a '+' here represents concurrent use and '–' represents the situations where nods are used alone. As an extension to this line of enquiry, plots 5 to 9 in Appendix 6.13 illustrate the basic distribution of each type of nod, **A** to **E** inclusive, mapping the points where they are used without concurrent spoken backchannels for each speaker/ conversation.

Speaker	Nod Type									
	A		B		C		D		E	
	+	-	+	-	+	-	+	-	+	-
S01FM.F	49	16	38	28	44	12	7	9	6	2
S01FM.M	14	47	4	24	2	8	0	1	0	0
S02MM.1	25	69	5	13	11	30	6	1	3	2
S02MM.2	85	41	95	34	38	4	11	0	33	9
S03MF.M	31	35	23	17	20	12	4	4	5	3
S03MF.F	57	146	8	22	38	36	1	0	2	1
S04MM.1	62	20	22	18	27	10	21	12	7	2
S04MM.2	83	61	41	34	16	10	3	0	1	3
S05MM.1	144	32	55	17	27	4	2	0	5	0
S05MM.2	78	59	6	5	25	12	0	1	1	0
S06FF.1	30	14	9	13	21	11	10	3	8	2
S06FF.2	23	35	59	155	16	11	12	12	19	5
	681	**575**	**365**	**380**	**285**	**160**	**77**	**43**	**90**	**29**

Figure 6.12: Frequencies of non-verbal backchannel behaviour, and its co-occurrence with spoken backchannels.

Figure 6.12 shows that in S01MF, S02MM and S05MM, type **A** were the nods most frequently used with spoken forms by all but one speaker, S02MM.2, where type **B** nods predominate. In S03MF, there was a greater proportion of type **A** nods used with spoken backchannels than without by

S03MF.F, whereas for S03MF.M more were used without than with spoken backchannels (54% are used with and 46% without for S03MF.M, 34% and 66% for S03MF.F - refer to Figure 6.12 for specific numerical frequencies that these percentages represent). S03MF.F also uses type **B** and **C** nods more frequently without spoken backchannels than with. This trend is reversed for S03MF.M.

S04MM.1 uses types **D** and **E** nods 15% and 5% of the total times that nods are used with spoken backchannels, whereas this is only 2% and 1%, respectively, for S04MM.2. This is also true for type **C** nods, with S04MM.1 using them far more often with spoken backchannels than S04MM.2, accounting for 22% and 11% of the respective total usage for these speakers. S04MM.2 uses nod types **A** and **B** more frequently with spoken backchannels than S04MM.1, together accounting for 89% of the total for S04MM.2, with 60% for type **A** and 29% for type **B,** and only 58% for S04MM.1, with 43% for type **A** nods and 15% for type **B** nods.

Again, for both speakers in S05MM, all nod types were used more frequently with, than without, concurrent spoken backchannels. The only exceptions to this are the type **D** nods performed by S05MM.2, in this case 100% are used without spoken backchannels. However since the frequency for this occurrence is 1, this result is not seen to be particularly significant. The rates with which S05MM.1 uses each type of nod with concurrent spoken forms is greater than those seen for S05MM.2. S05MM.1 uses 82% of type **A** nods, 76% of all type **B** nods, 87% of **C** and 100% of both **D** and **E** nods with spoken forms. For S05MM.2 these rates are 57%, 55%, 68%, 0% and 100% respectively.

Similarly, type **A** nods are also most commonly used with spoken backchannels by S06FF.1, 38% of the total concurrent nods and spoken forms for this speaker. These are followed by type **C** nods, 27% of instances. In contrast, S06FF.2 uses type **B** nods significantly more frequently with spoken backchannels than any other form, 46% of the total. She uses type **A** nods in only 18% of all instances, and types **C, D** and **E** on 12%, 9% and 15% of occasions.

Figure 6.12 also indicates that in five of the six videos, excluding S02MM, the supervisee uses both a greater net amount and personal proportion of type **A** nods without spoken backchannels, than the supervisor (refer to Appendices 6.9 and 6.10 for further details).

Additionally, throughout the complete dataset, nod types **C** and **D** are used slightly more frequently without spoken counterparts for those who speak more frequently in each supervision, although there is little difference for each supervisor and supervisee. Nod types **C** and **D** are often used with concurrent spoken backchannels by those who speak the least frequently in each dyad. This is true for 60% of the corpus data, however, results are not consistent from speaker to speaker.

In the majority of the cases, the speakers who use the least nods in each supervision, i.e. the most frequent speaker, uses the most type **B** nods, both with and without spoken backchannels. The only exception is video S03MF where the most frequent 'nodder', S03MF.F, uses type **B** nods in isolation more often than the other participant, although this is reversed for type **B** nods with accompanying spoken backchannels.

6.3.4.4. Spoken backchannels with nods- a focus on form

Figure 6.13 shows the types of head nods that most commonly co-occur with spoken backchannels of simple, double and complex structural forms.

Speaker	Spoken Backchannel Form						Speaker Total	Grand Total
	Simple		Double		Complex			
	+	-	+	-	+	-		
S01FM.F	23	17	16	1	112	81	250	287
S01FM.M	19	13	1	1	0	3	37	
S02MM.1	41	17	3	0	8	1	70	539
S02MM.2	303	71	53	7	23	12	469	
S03MF.M	57	41	20	4	20	18	160	292
S03MF.F	91	22	1	0	14	4	132	
S04MM.1	149	88	5	2	26	13	283	533
S04MM.2	147	82	7	3	6	5	250	
S05MM.1	236	79	4	0	20	3	342	487
S05MM.2	100	35	2	0	8	0	145	
S06FF.1	75	14	1	0	13	2	105	292
S06FF.2	130	50	0	0	3	4	187	
	1371	529	113	18	253	146		2430

Figure 6.13: Charting the frequencies of spoken backchannel forms and their co-occurrence with specific types of head nods.

The above table indicates that simple, double and complex forms of spoken backchannels are overall more likely to co-occur with than without a backchanneling head nod (compare the '+' and '-' columns). This is supported most convincingly for S05MM and S06FF where simple forms co-occur with nods more than 2.5 times more often then without This is true of 75%, 74%, 84% and 72% of instances where simple forms '+' nods are performed by S05MM.1, S05MM.2, S06FF.1 and S06FF.2. There is a fairly even amount of simple forms '+' and '–' nods for S01FM.M, with rates of 59% '+' and 41% '-',

but this is probably because the simple backchannels are used relatively infrequently by this speaker overall.

A similar situation exists for the double form backchannels. All speakers, except for S01FM.M and S06FF.2, use these forms more often with, than without, nods. Overall, double forms are used fairly infrequently, as S01FM.M uses them only once with and without nods and S06FF.2 does not use them at all. The speakers who do use double forms relatively often, with a frequency of >10, use these with nods at least twice as often as without. This includes S01FM.F, S02MM.2 and S03MF.M who each use 94%, 88% and 83% of all double forms with nods.

Finally, for the majority of these speakers, again complex spoken backchannels are more likely to be used with, than without, concurrent backchanneling nods. This is seen, most noticeably, in S05MM.2, S02MM.1, S05MM.1 and S06FF.1, all of whom use these forms at least 6 times more frequently with, than without, nods. This pattern is seen on 100%, 89%, 87% and 87% of all occasions when complex forms are used. The only exceptions are S01FM.M and S06FF.2 who use complex spoken backchannels most frequently without rather than with concurrent nods, in 100% and 57% of all respective instances, although the overall frequency of complex forms for these speakers is <10.

Figure 6.14 provides the most common functions that the top ten most frequent backchannel forms adopt (refer back to Figure 6.6),and details the rates at which these are used, with and without concurrent backchanneling nods (compare '+' with '-').

Lexical Form	Discourse Function								Total
	CON		CNV		ER		IR		
	+	-	+	-	+	-	+	-	
Mmm	577	216	0	0	0	0	0	0	793
Yeah	200	93	267	111	0	0	1	0	672
Yes	1	0	127	38	0	0	1	0	167
Right	0	1	24	16	2	0	49	24	116
Mhm	79	22	1	1	0	0	0	0	103
Okay	1	0	17	8	0	0	33	22	81
Mmm mmm	51	8	0	0	0	0	0	0	59
Yeah yeah	8	0	41	8	0	0	0	0	57
Sure	1	0	23	4	3	0	2	1	34
Uh hm	17	7	0	0	0	0	0	0	24
	935	347	501	185	5	0	86	47	
	1282		686		5		133		

Figure 6.14: The functions of the most commonly used backchannel forms, and the frequency with which they are used with and without backchanneling nods.

This indicates that each of the top ten backchannel forms are more likely to be used with, than without, a concurrent backchanneling nod, at a rate of at least 70% of the total for each. However, the exception is *right* and *okay* where this likelihood stands at 63% and 65%. The only spoken backchannel forms, with a frequency >1, which are more likely to be used alone rather than with concurrent nods are *erm*, *definitely* and the complex form phrases *yeah mm*, *yeah erm*, *right yeah yeah* and *well yeah* where 100%; 57%; 100%; 100%; 100% and 100% of their respective use is without concurrent nods (refer to Figure 6.14 for specific frequencies).

For both speakers in S01FM, *mmm* is more likely to be used alone, rather than with concurrent nods, a characteristic not seen throughout the other

videos in this corpus, except for S03MF.M, who only uses *mmm* 3 times, twice without nods and once with.

In terms of S02MM, the analysis revealed that all of the forms featured in the top ten, are more likely to co-occur with nods than to be used in isolation. This is true of both speakers. This characteristic is also seen for all speakers in S03MF and S04MM. The only exceptions to this are *right yeah yeah* for S03MF.M and *erm* for S03MF.F, which are used without nods for 100% of all instances. *Oh yeah*, spoken by S03MF.M, *mmm* and *no* spoken by S03MF.F are all used in equal amounts with and without nods.

In S04MM, the backchannels *no*, *yeah mmm* and *okay* are used 100% of the time without nods, whereas *right,* in the case of S04MM.1, is used equally with and without.

In S05MM and S06FF, nearly all of these forms are proportionally more likely to co-occur with nods than to be used in isolation. The only instances where this is not the case is in S05MM, with the use of *right,* by both speakers, where in 75% and 50% of instances this is used without concurrent nods by S05MM.1 and S05MM.2 respectively. Additionally, the results show that the most common spoken backchanneling forms used in S06FF, i.e. those with a frequency of >2, are more likely to be co-occur with backchanneling nods than without, with the exception of the backchannel *yeah true* which is used an equal amount of times by S06FF.2 for each, 1 with and 1 without.

Figure 6.15 shows the relationship between the functions of these spoken backchannels, and the type of nod with which they co-occur.

Spoken Form	Discourse Function	Concurrent Head Nod Type				
		A	B	C	D	E
Mmm	CON	266	186	46	23	56
Yeah	CON	77	61	35	13	14
	CNV	104	59	61	18	25
	IR	1	0	0	0	0
Yes	CON	0	1	0	0	0
	CNV	55	22	24	11	15
	IR	0	0	0	1	0
Right	CNV	13	3	8	0	0
	ER	0	0	2	0	0
	IR	16	5	22	3	3
Mhm	CON	15	37	7	8	12
	CNV	0	0	0	1	0
Okay	CON	0	1	0	0	0
	CNV	11	0	4	1	1
	IR	13	4	11	3	2
Mmm mmm	CON	9	21	4	2	15
Yeah yeah	CON	0	3	3	1	1
	CNV	15	13	6	4	3
Sure	CON	0	0	0	1	0
	CNV	9	4	5	0	5
	ER	2	1	0	0	0
	IR	2	0	0	0	0
Uh hm	CON	8	3	2	0	4
		616	424	240	90	156

Figure 6.15: The relationship between discourse function and concurrent nod type (for the top 10 most frequent spoken backchannel forms).

The figure indicates that nod types **A** and **B** are most frequently used with these top-ten most frequently used forms, amounting to 40% and 28% of the total. The only exceptions are *mhm* and *mmm mmm,* where the most common concurrent nod used, in both instances, is type **B,** in 36% of the total. Type **D** nods are the least frequently used with these forms.

Figure 6.15 also shows that for these top-ten forms, those that adopt the CON function most commonly co-occur with type **A** nods, as seen on 40% of the total uses of spoken backchannels functioning as CON tokens. These are closely followed by type **B** nods, in 33% of all instances. For backchannels

functioning as CNV tokens, type **A** nods are used on 41% of all instances while types **C** and **B** are used 22% and 20%.

Spoken backchannels adopting an IR function in the corpus are used with type **C** nods in 38% of all instances, closely followed by type **A** nods, with 37%. Finally, ER tokens, around 40%, are used with type **A** and 40% as type **C** nods, although since the total frequency for this function is 5, the significance of this result is negligible.

In the majority of supervisions examined in the corpus, spoken backchannels functioning as CON, CNV, ER and IR tokens are most often used with rather than without concurrent nods, supporting premise 8. This is true for 73% of the total for CON and CNV tokens, 100% for IR tokens and 65% for IR tokens.

6.3.4.5. Spoken backchannels with nods- a focus on function

Figure 6.16, overleaf, provides a breakdown of the frequency with which *all* spoken backchannels, and associated discourse functions, are used with and without concurrent nods. Refer to Appendices 6.9 and 6.10 for a breakdown of these results.

Figure 6.16 illustrates that for 100% of the speakers, spoken backchannels that adopt CON and CNV functions are used either the same amount, or more frequently with, than without, backchanneling nods. The only exception to this is S01FM.M who uses CON tokens the same amount of times with and without concurrent nods. To a certain extent, this relationship is also seen for the IR tokens, although there are more exceptions, as S01MF.M, S01MF.F

and S04MM.2 use these tokens more frequently without rather than with concurrent nod although the frequency of use is relatively small.

Speaker	Discourse Function								Total
	CON		CNV		ER		IR		
	+	-	+	-	+	-	+	-	
S01FM.F	86	77	59	9	3	6	3	7	250
S01FM.M	7	7	11	7	1	1	1	2	37
S02MM.1	10	3	23	10	8	2	11	3	70
S02MM.2	286	59	76	17	12	12	5	2	469
S03MF.M	30	21	36	14	7	10	24	18	160
S03MF.F	39	13	52	9	5	3	10	1	132
S04MM.1	100	57	55	34	16	8	9	4	283
S04MM.2	93	39	60	42	2	2	5	7	350
S05MM.1	152	45	89	31	12	4	7	2	342
S05MM.2	72	25	30	8	5	1	3	1	145
S06FF.1	11	6	52	4	6	1	20	5	105
S06FF.2	78	19	38	23	3	1	14	11	187
	964	371	581	208	80	51	112	63	2430

Figure 6.16: Frequency with which spoken backchannels are used with and without concurrent nods across the five-hour corpus.

Figure 6.16 indicates that for S01MF.M, S01MF.F, S02MM.2, S03MF.M and S04MM.2, those spoken backchannels functioning as ER tokens are used either the same amount of times or more frequently without rather than with concurrent nods, accounting for 66%, 50%, 50%, 59% and 50% of the total number of ER tokens used by each. Since the ER tokens are the least frequently used overall, the difference between those used with and without nods is smaller than for the other three functions.

In short, the speakers who most frequently use spoken backchannels with concurrent nods use a higher proportion of these as CON and CNV tokens. This is true of S02MM.2, S05MM.1, S01FM.F, S04MM.1 and S06FF.2, where

participants speak less in each supervision video. Those who use spoken and non-verbal backchannels concurrently, at a less frequent rate than the other speaker in the dyad (refer back to Figure 6.11 for details), use a higher proportion of such as IR tokens, as for S05MM.2, S03MF.M, S06FF.1, S02MM.1 and S01FM.M.

As a final observation, it should be noted that 5 out of 6 of the supervisors use a higher proportion of ER and IR tokens with concurrent nods than their supervisee, from the total number of concurrent spoken and non-verbal backchannels for the given speaker. IR tokens are also used more frequently by the participants who speak more frequently in each of the videos. See Appendices 6.7 and 6.8 for a detailed summary of these results.

6.3.4.6. The relationship between lexical function and nod type

Figure 6.17, overleaf, provides a detailed breakdown of the frequencies of the individual nods that are used concurrently with the spoken backchannels, listing the functions adopted by the spoken backchannels and the type of concurrent nods (i.e. detailing spoken to non-verbal backchannels). Also refer to section 4 of Appendices 6.1 to 6.6 for a breakdown of these results.

Figure 6.17 indicates when S01FM.F uses types **A** and **B** nods, these are most likely to co-occur with spoken backchannels functioning as CON tokens , as seen for 34% and 36% of all concurrent nods and CON (see figure for specific frequencies). Whereas, type **C** are more likely to co-occur with those functioning as CNV tokens, as seen with 31% of all nods and concurrent CNV. There is only a slight difference between the use of nod types **D** and **E** and concurrent CON and CNV tokens, each representing <10% of the

respective totals. In addition, both ER and IR tokens most frequently co-occur with nods of type **C**, although since the frequency for each of these stands at only 2, so far from conclusive.

Speaker	Discourse Function (Colour) and Concurrent Nod Type (Letter)																			
	A	**A**	**A**	**A**	**B**	**B**	**B**	**B**	**C**	**C**	**C**	**C**	**D**	**D**	**D**	**D**	**E**	**E**	**E**	**E**
S01FM.F	29	18	1	1	31	10	0	0	18	22	2	2	3	5	0	0	5	4	0	0
S01FM.M	5	8	0	1	2	1	1	0	0	2	0	0	0	0	0	0	0	0	0	0
S02MM.1	6	12	4	3	1	2	1	2	0	5	2	4	3	1	0	2	0	3	1	0
S02MM.2	65	18	2	1	115	23	2	0	21	15	4	1	15	1	0	0	70	19	4	3
S03MF.M	7	13	2	9	13	15	1	4	5	5	0	10	1	2	1	0	4	1	3	1
S03MF.F	23	27	2	5	2	3	2	1	11	22	1	4	1	0	0	0	2	0	0	0
S04MM.1	39	14	6	3	21	10	5	0	15	8	2	2	15	16	2	4	10	7	1	0
S04MM.2	45	37	0	1	40	13	2	0	6	6	0	4	2	2	0	0	0	2	0	0
S05MM.1	91	45	5	3	53	20	4	1	7	18	1	1	0	2	0	1	1	4	2	1
S05MM.2	57	17	3	1	4	2	0	0	11	10	2	2	0	0	0	0	0	1	0	0
S06FF.1	7	14	1	8	0	6	2	2	1	13	1	6	3	7	1	2	0	12	1	2
S06FF.2	12	7	0	4	39	15	3	2	6	5	0	5	8	3	0	2	13	8	0	1
	386	230	26	40	321	120	23	12	101	131	15	41	51	39	4	11	105	61	12	8
	682				476				288				105				186			

Figure 6.17: Exploring the relationships between the spoken functions and nod types of concurrent spoken and non-verbal backchannels, across the five-hour corpus.

Of the 20 concurrent nods and spoken backchannels used by S01FM.M, 71% of those spoken forms functioning as CON tokens are used with nod type **A** and 72%, functioning as CNV tokens are used with those nods of type **B**. There were no recorded instances of nod types **D** or **E** co-occurring with spoken backchannels for this speaker.

S02MM.2 uses CON and CNV tokens most frequently with type **B** nods, amounting to 40% and 30% of all CON and CNV tokens used by this speaker. For both functions, this is followed by type **E** and **A** nods. Of the type **A** nods

used by S02MM.1, 48% are used as CNV tokens, whereas only 24% are used as CON tokens. However, for this speaker type **A** nods prove to be the most frequently used nods with spoken backchannels adopting all discourse functions apart from IR tokens. Type **A** nods are second most frequent nods with spoken backchannel forms.

S03MF.M uses type **B** nods most frequently with backchannels functioning as CON tokens, amounting to 26% of all these nods types used by this speaker. By comparison, S03MF.F uses 75% of her CON tokens with nods, 59% of which co-occur with type **A** and 28% with type **C** nods.

Figure 6.17 also shows that S04MM.1 uses 56% of all of his type **E** nods with CON tokens. Whereas S04MM.2 does not use any type **E** nods with backchannels functioning as CON tokens. Additionally, S04MM.2 uses almost twice the number of type **B** nods with CON tokens than those used by S04MM.1. S04MM.2 also uses type **A** nods with CON and CNV tokens more frequently than S04MM.1, at 99% and 85% of all type **A** nods used. Whereas S04MM.1 uses far more type **C** and **D** nods with CON and CNV tokens than S04MM.2, although S04MM.1 performs more type **C** and **D** nods overall.

The results also indicate that S05MM.1 uses CNV tokens most frequently with concurrent type **C** nods, followed by CON tokens with type **C** nods. This pattern is the reverse for S05MM.2. Type **A** nods are most commonly used with ER tokens for both speakers in this video, at a rate of 42% and 60% of all ER tokens used by S05MM.1 and S05MM.2 respectively. These were followed by types **B, E, C** and **D** nods for S05MM.1 and type **C** for S05MM.2. There are no spoken backchannels functioning as ER tokens co-occurring with nod types **B**, **D** and **E** for this speaker. Type **A** nods are also used most

frequently with IR tokens for S05MM.1, although type **C** nods were most common for S05MM.2.

At least 60% of the type **A** nods used with spoken backchannels in S05MM are used with CON tokens. This rate stands at 63% for S05MM.1 and 73% for S05MM.2 (refer to Section 4 of Appendix 6.5). Of the 78 type **B** nods concurrently used with spoken backchannels by S05MM.1, the most commonly associated functions are CON tokens, at 73%, followed by CNV tokens at 26%. This pattern is also matched by S05MM.2.

In terms of the final supervision, S06FF, the results indicate that S06FF.1 uses CNV, CON and IR tokens most frequently with type **A** nods, as seen on 27%, 64% and 40% of the total that such functions are used. She also uses ER tokens most frequently with type **B** nods; on 33% of occasions where spoken backchannels are seen to function as ER tokens. CNV tokens are also readily used with type **C** and **E** nods, with frequencies of 13 and 12 respectively for this speaker, 25% and 23% of the total, each of which are used nearly as frequently as the amount used with type **A** nods.

On the other hand, S06FF.2 uses far more concurrent spoken backchannels and nods than S06FF.1, using type **B** nods most prevalently amounting to 44% of all those used for this speaker. Type **B** nods most commonly co-occur with CON and CNV and ER tokens for S06FF.2, and IR tokens are most frequently used with type **C** nods, although this is closely followed by type **A** nods. These account for 50%, 39%, 100%, 36% and 29% of the occasions where these nods are used with the respective functions.

Overall, then, Figure 6.17 illustrates that for 8 of 12 the speakers, backchannels functioning as CON tokens are used with type **A** nods more

frequently than other nods, amounting to 40% of the total value of CON tokens used with nods. Whereas, for S01MF.M, S02MM.2, S03MF.M and S06FF.2 type **A** nods are the second most commonly used with CON tokens. This is true for all speakers except for S02MM.2 who uses type **E** nods second most frequently with CON tokens, as type **B** nods are the most common, representing 33% of this total. 7 of the 12 speakers use IR tokens with type **A** nods most frequently, while 50% of the speakers are shown to use CNV tokens with type **A** nods and/or ER tokens with type **C** nods most frequently, at 40% and 37% of the totals for each of these.

Type **D** nods are shown to co-occur less frequently with CON and CNV tokens than the other nod types, amounting to only 5% and 7% of the total throughout all supervision videos, for each respective function. These are followed by nod types **C** and **E,** for CON tokens, and types **E** and **B** nod, for CNV tokens. ER tokens are least frequently used with type **D** nods, followed by type **E** nods, each comprising nods 5% and 15% of the total. Whereas for IR tokens, type **E** nods are the least commonly used with concurrent spoken forms, closely followed by type **D** nods, comprising 7% and 10% of each respective total.

In terms of the proportion with which these nods are used with each spoken function, Figure 6.17 highlights that type **A** nods are used most often with CON and CNV tokens, amounting to 40% of the total for each. Type **B** nods are also used with CON tokens at a higher rate than for other tokens, that is, 33% of the total. IR tokens co-occur with type **C** and **D** nods most frequently, at 37% and 10% of all IR tokens used, and type **E** nods most

commonly co-occur with backchannels functioning as ER tokens, at 15% of all these tokens.

Furthermore, the supervisors use concurrent type **A** nods with ER tokens more frequently than the supervisees, from the total frequency of nod type **A** use for the given speaker. This is true of 5 out of 6 of the supervisions, except for S05MM. This pattern also proves true for the most frequent speakers in these videos, with the exception of S01FM. Similarly, the supervisors use both type **B** nods with IR tokens, and type **E** nods with ER tokens, more frequently than the supervisees. This was seen in at least 5 of the 6 supervisions, for each of these conditions.

6.3.4.7. Summary

In short, spoken and non-verbal backchannels are seen to co-occur on many occasions in this corpus. The basic patterns of this co-occurrence are summarised below:

M. Of the 1498 different nods seen in this corpus, 1187 of them co-occurred with spoken backchannels, and of the 2430 spoken backchannels used, 1737 of them co-occurred with a backchanneling nod. In other words for >70% of the times that a spoken or non-verbal backchannel is used, it co-occurs with a non-verbal/ spoken form. This finding supports premise 8.

N. Each of the top 10 most frequently used lexical forms of the spoken backchannels were *more likely to be used with concurrent backchanneling nods* than to be used in isolation.

O. All of the speakers examined here use CON tokens *with backchanneling nods as frequently or more frequently than without accompanying nods*. This is also true, in all but one instance, for IR tokens. Participants are shown to use ER tokens more frequently without accompanying nods on all occasions. This is true for CNV tokens on all but one occasion.

P. There is no real relationship between nods of a longer duration (types **B**, **D** and **E**) and their frequency of use with/without concurrent spoken backchannels.

Q. The type of backchanneling nods used in the corpus relate closely to the lexical structure and discursive function of concurrent spoken backchannels. More affective and 'complex' forms of spoken backchannels are more likely to be used with head nods of a complex structure, so of a longer duration and/or variable intensity, that is, types **B, D** and **E**. Whereas simple form nods, types **A** and **C** for example, co-occur with simple structural forms of spoken backchannel behaviour. This finding partially supports premise 9.

6.3.5. Backchanneling in context

6.3.5.1. Overview

As identified in Section 6.3.4.1, there are many instances in the corpus where a backchanneling nod is used across multiple turn boundaries and, thus, at

specific locations across more than one spoken backchannel. Therefore, it is now appropriate to explore the simple patterns of positioning of backchanneling head nods, in the context of the remainder of the conversation; i.e. investigating points where nods precede or follow a speakers' turn. The question below will be addressed as part of this line of enquiry:

10-   Spoken and non-verbal backchannels are often used collaboratively in talk, and are shown to cluster and operate in context: within and across turn boundaries.

6.3.5.2. Backchanneling across turns

To aid in this line of investigation, Appendices 6.11 and 6.12 chart the types and frequencies of the head nods that co-occur with spoken backchannels across turn boundaries in the corpus (refer to Section 6.4.3.1 for more details). Again, this includes those nods which are either used with multiple spoken backchannel forms and/or nods that precede or follow the use of a single spoken form.

Both appendices demonstrate whether the nods are performed before the spoken form is verbalised, and/or whether they continue after it. These are labelled as 'bf', 'af' and a&b', respectively in these appendices. Appendix 6.11 provides a breakdown for each individual video, and Appendix 6.12 combines the results for ease of reference. Specific details of these, for each video, can be seen in Section 6 of Appendices 6.1 to 6.6 for S01FM to S06FF inclusive.

Overall, the results indicate that type **B** nods that co-occur with spoken backchannels are the most likely to be used across turns for both speakers in this video, with 33 instances of this occurring with S01FM.F and 2 instances with S01FM.M. This is true for 33 of 40 (83%), and 2 from 4 (50%) of the occasions where type **B** nods and spoken backchannels are concurrently used by S01MF.F and S01MF.F. In the majority of these instances, nod type **B** precedes the actual utterance of the spoken backchannel. This occurs 27 times out of the combined total of 35.

Furthermore, Section 1b of this appendix illustrates that S01MF.F uses nod types **D** and **E** each on 7 occasions across turns, 6 and 4 times respectively before the spoken backchannels, as might be expected for nods of a long duration. This speaker also uses 6 type **C** nods across turns, 5 of which begin prior to the verbalisation of the spoken backchannel.

The appendix highlights that derivations of the backchannel *mmm* and *yeah*, including *mmm mmm*, *yeah yeah,* are frequently used across turns for S01FM.F. Moreover, it shows that it is likely that the co-occurring nod starts before these spoken forms are uttered, as seen in 73% of all instances. Overall, backchanneling nods are used across turns for 37%, 56 from 151, of all instances where concurrent nods and spoken forms are used by this S01MF.F, a proportion that is much greater than that seen for S01FM.M who uses 4 out of 20 (20%) of concurrent backchannels in this way.

The simple forms *mmm* and *yeah,* functioning as CON, are most frequently used with nods across turns for S01FM.M, with the nod preceding the spoken form in every instance. Of all forms of spoken backchannels, the forms *mmm,* functioning as a CON, and *yeah* are commonly used across

turns. Overall, in 60% of all of the instances where S01FM.F uses *yeah* with a concurrent nod, the nod starts before the lexical item is actually spoken. This is true for 12 of the 20 instances presented; across both CON and CNV functions. This stands at 100% for S01FM.M, that is, with a frequency of 1.

As with S01FM, there are many occasions where a backchanneling nod is used over two or more turns and spoken backchannel forms in S02MM, although the vast majority of these cases are performed by S02MM.2. Sections 2a and 2b of Appendix 6.11 provide details of backchannels that occur across turn boundaries for S02MM. Here 19 and 258 spoken backchannels are shown to co-occur with nods stretching beyond the backchanneling turn for each of these respective speakers, amounting to 37% and 68% respectively of each of the total concurrent spoken and non-verbal backchannels performed by them.

The most common spoken backchannel form and discursive function used in this way by S02MM.2 is *mmm* used as a CON, where in 64 cases the nod is initiated prior to the verbalisation, 40 instances where it continues after it, and 45 cases involving a combination of these, 73% of the total for this spoken form. In these cases, *mmm* is most commonly used with a type **B** nod. As with S01FM, *yeah* is commonly used across turns in this supervision, particularly for S02MM.2. There are 27 instances where this is seen in the data, with 19 functioning as CNV tokens and 8 functioning as a CON, amounting to 66% of the total that this token is used with a concurrent nod by this speaker. The table also indicates that type **B** and type **E** nods are most frequently used across turns for both of the speakers in S02MM.

Sections 3a and 3b of Appendix 6.11 illustrate the forms and functions of spoken backchannels that most commonly co-occur with nods used across turns for S03MF. Unsurprisingly, nod type **B** is most frequently used in this way, usually with *yeah* functioning as a CON. This occurs in 21 instances for these two speakers.

As with S02MM, there is a range of different spoken forms, with different discourse functions, that are used with nods across turns for S03MF (see Sections 3a and 3b of Appendix 6.11). Both speakers have a tendency to start these nods before the spoken backchannel is uttered, using type **B** nods in the majority of cases, that is >50% of the total for both speakers, 10 out of 19 times for S03MF.M and 13 out of 24 for S03MF.F.

In S04MM (see Sections 4a and 4b of Appendix 6.11), the spoken forms most commonly used across turns by both participants, are *yeah*, and *mmm*, all of which most prevalently co-occur with either type **B, D** or type **E** nods. In this supervision, both *yeah* and *mmm,* functioning both as CNV and CON tokens for *yeah* and CON for *mmm*, and co-occurring with type **B** nods, are generally initiated before the backchannel is uttered. Although for S04MM.1, there is a fairly even balance of nods that precede or follow the spoken backchannel, whereas for S04MM.2, such nods most frequently precede the spoken backchannel.

This pattern of behaviour is also seen for S05MM (refer to Sections 5a and 5b of Appendix 6.11). Here type **B** nods; co-occurring with *mmm* and *yeah* backchannels, and functioning as CON and CNV tokens, are most commonly used across turns. The only difference between the speakers in S05MM is that S05MM.1 uses such tokens with nods that generally start *with* the

verbalisation of the backchannel and continue after it. This is true of 24 instances, amounting to 38% of the total *mmm* + nod combination for this speaker. However, there are a further 16 cases where the nod precedes the backchannel for this speaker, although there are no instances of this for S05MM.2, although the total frequency of nods across turns is much lower. The backchannel *yes,* rather than the form *yeah,* functioning as CNV, is also used in this way by S05MM.1. That is*,* with the nod starting at the same time as the verbalisation, and continuing afterwards.

Again, type **B** and **E** nods are most commonly used across turns for both speakers in S0F66. The majority are initiated before the concurrent spoken backchannel, mainly occurring with *yeah* and *mmm,* but in the case of S06FF.2, *yes*, functioning as a CNV (refer to sections 6a and 6b of Appendix 6.11). In addition, the results reveal that S06FF.2 uses *mhm* functioning as a CON with concurrent type **B** nods that follows the verbalisation in 27 instances, i.e. on 41% of all occasions where it is used with a nod.

Overall, Appendix 6.12 shows that of the 1737 spoken backchannels that co-occur with nods in this corpus, 767 (44%) are used with nods that either precede and/or follow the spoken backchannel. In 30% of these cases, the nod, most often type **B**, precedes the verbalisation, while in 15% the type **B** nod will continue after the spoken form has been uttered.

In addition, 492 CON, 213 CNV, 29 ER tokens and 33 IR tokens are used with nods across turns in some way, amounting to 51%, 37%, 36% and 29% of the successive total concurrent spoken backchannel and nods used in the corpus. Of these nod types **B, B, B** and **D** are most frequently used, with the nod starting before the spoken backchannel in the majority of these instances.

In terms of spoken backchannels functioning as CON tokens, the forms that are most frequently used across turns include *mmm*, *yeah*, *mhm* and *mmm mmm* each with a total of 280, 91, 59, 36 instances. In each case, the majority of these spoken forms are used with type **B** nods that start before the utterance, representing 48%, 46%, 75% and 71% of the total amount these spoken backchannels are used with nods in the corpus. The most common CNV tokens used in this way are *yeah*, *yeah yeah*, *yes* and *sure* with frequencies of 101, 17, 51 and 11 respectively, i.e. 38%, 40%, 40%, 28% of the total these spoken backchanneling forms are used with concurrent nods. The most common ER token, with a frequency of ≥3, used across turn boundaries is *definitely*, with 3 instances, that is, on 100% of the times this spoken form is used, it co-occurs with nods in the corpus. The most common spoken forms functioning as IR tokens used in this way are *right* and *okay,* with 15 and 9 occurrences, equating to 52% and 27% of their respective totals.

6.3.5.3. Sequences of backchannel use

As an extension to the exploration, an examination of the types of co-occurring spoken and non-verbal backchannel behaviour which successively precede or follow each other across the corpus can be undertaken. For this, it is necessary to concentrate on the sequential behaviour of the most common concurrent spoken and non-verbal backchannels. Overall, the most common spoken/non-verbal backchannel combinations are a type **A** nod which co-occurs either with a CON or CNV token, and type **B** nods that are used with

CON and CNV tokens. Therefore, type **A** and **B** nods have been concentrated on, given that these have proved to be most frequently used in the corpus.

Despite the frequency of type **A** nods, the most common sequence of concurrent backchannel behaviour, i.e. with 2 spoken/non-verbal backchannel combinations in succession, is actually for a nod of type **B**, co-occurring with a spoken backchannel functioning as a CON, followed with another type **B** nod, co-occurring with a CON spoken backchannel. This occurs in 105 instances throughout the five-hour corpus. Type **A** nods co-occurring with CON backchannels are followed, in 84 instances, by another type **A** nod with a concurrent CON backchannel. These nods are followed by type **A** nods with CNV tokens 47 times, equalling the frequency with which type **A** nods with CON tokens are followed by type **B** nods with CON. Figure 6.18 details the top ten most frequent sequences of backchannel use, across type **A** and **B** nods.

Initial		Followed by		
Nod Type	Spoken BC function	Nod Type	Spoken BC function	Frequency
B	CON	B	CON	105
A	CON	A	CON	84
A	CON	A	CNV	47
A	CON	B	CON	47
B	CON	A	CON	43
A	CNV	A	CNV	39
A	CNV	A	CON	34
B	CON	B	CNV	28
B	CNV	B	CON	24
A	CON	B	CNV	18

Figure 6.18: Exploring sequences of concurrent spoken and non-verbal backchannels.

Of the patterns of backchanneling nod usage depicted in Figure 6.18, 48, (46%) are enacted by S02MM.2, while S06FF.2 and S04MM.2 both use this combination of concurrent backchannels on 10 occasions. These amount to around 10% of the total, see Figure 6.19 for further details of sequences of backchannel use across each speaker. It is important to note that although this enquiry focused on 'sequences' of behaviour, these sequences do not directly follow each other in the context of the conversation. Instead, these indicate backchannels are likely to follow others over time, irrespective of whether they are used in subsequent turns in the discourse.

Speaker	Initial nod/ function	Followed by	Freq.	Initial nod/ function	Followed by	Freq.
S01FM.F	A, CON	B, CON	6	B, CON	B, CON	5
S01FM.M	A, CON	A, CNV	2	A, CNV	A, CON	2
S02MM.1	A, CNV	A, CNV	2	A, CON	A, CNV	1
S02MM.2	B, CON	B, CON	48	B, CON	B, CNV	15
S03MF.M	B, CON	B, CON	7	B, CNV	B, CNV	6
S03MF.F	A, CON	A, CNV	8	A, CON	A, CON	6
S04MM.1	A, CON	A, CON	9	B, CON	B, CON	9
S04MM.2	B, CON	B, CON	10	B, CON	A, CON	9
S05MM.1	A, CON	A, CON	28	B, CON	B, CON	16
S05MM.2	A, CON	A, CON	15	A, CNV	A, CNV	6
S06FF.1	A, CNV	A, CNV	2	All others have freq. <1		
S06FF.2	B, CON	B, CON	10	B, CNV	B, CON	5

Figure 6.19: Exploring the patterns-of-use of the most common sequences of concurrent spoken and non-verbal backchannels used by each speaker.

6.3.5.4. The lexical 'context' of backchannel use

The lexical 'context' in which backchannels are used across the complete corpus can also be examined. That is, the patterns of lexis that are often used prior to and/or following the use of spoken and/or non-verbal backchannels.

In order to conduct this enquiry, the approximate positions where both spoken backchannels are used and where backchanneling nods start, when not used with concurrent spoken forms, have been encoded across the complete corpus. Thus, with each simple, double and complex spoken form re-classified as a single spoken unit. Subsequently, by using the Collocate tool in Wordsmith it was possible to search for the following:

- Individual collocates and clusters of words that are frequently used in close proximity of backchanneling behaviours. See files 6.7, 6.9, 6.10, 6.12, 6.13 and 6.15 of the data disk for raw results of these outputs.

- Concordance outputs from the immediate lexical co-text of specific backchannels. See files 6.8, 6.11 and 6.14 of the data disk for raw results.

In terms of collocates, the raw outputs indicate that the majority of both spoken and non-verbal backchannels are in close proximity to grammatical lexemes, that is, function words with little lexical meaning rather than content words, i.e. words with a specific lexical content.

Of the grammatical, function words used, conventionalised forms of deictic markers are particularly frequent. This includes the use of personal pronouns (including *you*, it, them, he, *I*, *we*), determiners (including *the*, *a*), and demonstrative directive adverbs (including *this* and *that*), all of which commonly feature in the most common collocates, both across instances of spoken and non-verbal backchannel use.

If the specific positions of collocates, in relation to the use of backchannels, are examined, it can be revealed that while this pattern of the use of function words is generally true across the results, an interesting exception to this is seen with the lexical items that are used at positions R1, directly following the point at which the given form of non-verbal backchannel, specifically, has been performed. As seen in Figure 6.20, there is in fact a cluster of content words used at R1, following the use of head nods. Also see file 6.7 on the data disk for a raw, unedited version of these results.

Rank	Lexical item	Rank	Lexical item	Rank	Lexical item
1	keep	11	useful	21	put
2	theoretical	12	chapters	22	briefly
3	nineteenth	13	references	23	framework
4	they've	14	understanding	24	spaces
5	perspective	15	class	25	other
6	literature	16	even	26	listening
7	perhaps	17	critical	27	getting
8	metonymy	18	verbs	28	language
9	moment	19	certainly	29	come
10	literary	20	body	30	big

Figure 6.20: Lexical collocates that most frequently follow the use of backchanneling nods in the corpus (i.e. located at position R1).

In this figure nouns (including *perspective*, *literature*, *metonomy*, *moment*, *chapters*, *references*, *class*, *verbs*, *framework, listening, language* and *body*) are predominantly used at the onset of a backchanneling nod, followed by adjectives (including *theoretical*, *nineteenth*, *literary*, *useful*, *critical*, *certainly*, *briefly* and *big*), then some verbs and adverbs, while function words only feature in a couple of instances among this top 30 of most frequently collocates.

It is important to note that many of the nouns, and some of the adjectives, used here are somewhat context and domain specific, insofar as the frequent use of these lexemes probably results from these recordings being taken from academic supervisions in the department of English in a university. Therefore it is unlikely that such lexemes will prove as frequent across other speakers in other discursive environments. However, a comparative study of this would need to be undertaken to support this claim.

The predominant use of content words at R1 contrasts with those used at L1, that is, prior to the start of a backchanneling head nod. This is detailed in Figure 6.21.

Rank	Lexical item	Rank	Lexical item	Rank	Lexical item
1	the	11	you	21	about
2	of	12	on	22	at
3	to	13	in	23	chapter
4	a	14	is	24	you're
5	and	15	or	25	just
6	yeah	16	think	26	that's
7	that	17	with	27	are
8	erm	18	what	28	this
9	pause	19	well	29	it's
10	it	20	have	30	there

Figure 6.21: Lexical collocates that most frequently precede the use of backchanneling nods in the corpus (i.e. located at position L1).

This figure indicates that only one noun, *chapter*, is used before the nod, and only a few verbs, including *is*, *think*, *have* and *are*, and adverbs *just*, *there* and *well*, are also used here. The majority of terms in this location are again grammatical function words such as prepositions, including *on*, *in*, *about*, *at*, *with*, *of* and *to*, and determiners, such as including *the*, *a*, *that*, *what* and *this*.

Again, this general pattern of results is also seen at the majority of other locations, from L5 to R5, around the use of backchanneling nods. This is also occurs in the lexical environment of spoken backchannels, from L5 to R5.

As a point of comparison, Figure 6.22 details the lexemes which most frequently precede the use of spoken backchannels in the corpus, both with and without concurrent backchanneling nods.

Without concurrent nods				With concurrent nods			
Rank	Lexical item	Rank	Lexical item	Rank	Lexical item	Rank	Lexical item
1	it	16	with	1	it	16	space
2	that	17	you	2	that	17	do
3	of	18	the	3	and	18	things
4	there	19	before	4	erm	19	words
5	erm	20	adjectives	5	mean	20	but
6	to	21	mean	6	to	21	different
7	think	22	data	7	way	22	well
8	but	23	chapter	8	there	23	about
9	know	24	space	9	in	24	on
10	yeah	25	not	10	of	25	data
11	do	26	well	11	know	26	so
12	be	27	this	12	you	27	or
13	because	28	somebody	13	chapter	28	is
14	them	29	way	14	them	29	work
15	so	30	actually	15	yeah	30	before

Figure 6.22: Lexical collocates that most frequently precede the use of spoken backchannels (with)out concurrent nods (located at position L1).

Figure 6.22 highlights that, as with backchanneling nods, grammatical, function words are again used frequently at L1, with *it* and *that* proving to be the most commonly used prior to spoken backchannels, with and without nods. Further to this, as shown in Figure 6.23, a similar pattern is seen with

collocates of spoken forms at position R1, that is, directly after the use of spoken forms, both with and without concurrent nods:

Without concurrent nods				With concurrent nods			
Rank	Lexical item	Rank	Lexical item	Rank	Lexical item	Rank	Lexical item
1	oh	16	there's	1	mm	16	are
2	now	17	week	2	uh	17	imagine
3	when	18	which	3	thought	18	haven't
4	other	19	but	4	course	19	one
5	language	20	er	5	that's	20	study
6	um	21	up	6	themes	21	wouldn't
7	I	22	more	7	once	22	I'd
8	probably	23	have	8	thinking	23	doing
9	road	24	is	9	absolutely	24	with
10	cos	25	on	10	response	25	individual
11	here	26	okay	11	he	26	postcards
12	interesting	27	laughs	12	something	27	process
13	go	28	another	13	healthcare	28	no
14	use	29	stuff	14	even	29	review
15	if	30	into	15	needs	30	we've

Figure 6.23: Lexical collocates that most frequently follow the use of spoken backchannels (with)out concurrent nods (located at position R1).

Again spoken backchannels used in isolation are generally followed by function words, although the ratio here is only 16: 14, and spoken forms with concurrent nods prove to use a larger amount of content words at R1, with 13: 17 function to content words used here. This includes a variety of nouns (examples include *course*, *themes*, *healthcare*, *study*, *individual*, *postcard* and *process*), adverbs (including *even* and *no*), verbs (*thought*, *thinking*, *needs*, *imagine* and *doing*) and adjectives (*absolutely*).

However, overall, while there is greater use of content words at R1 for spoken backchannels with and without concurrent nods, in comparison to L1, the dominance of these word forms at this position is not as significant as that

seen in Figure 6.20. So while non-verbal forms remarkably differed from L1 to R1, there is generally a more stable pattern of collocates for spoken backchannels across these two positions. However, the other positions, from L5 to R5 are more evenly balanced, as seen in data disk files 6.9, 6.12 and 6.15.

This is interesting because, as identified in Chapter 2, both non-verbal and spoken backchannels were thought to commonly appear 'after nouns, verbs and adverbs' (Blache et al., 2008: 114). Whereas, in this corpus non-verbal backchannels, those which are used in isolation, are frequently used directly after prepositions, pronouns and determiners, preceding rather than following nouns, verbs and adverbs.

Finally, in terms of the specific clusters of words used in the immediate discursive environment of backchannels, Appendices 6.16-6.18, indicate that there is no real difference between the close lexical co-text where spoken and non-verbal backchannels are used. What is interesting to note, however, is the frequent use of interpersonal discourse markers across the most frequent lexical clusters. This includes uses of the phrases *kind of* and *sort of*, and derivations of the expressions *do you know what I mean*, *you see what I mean*. Discourse markers, also known by a multitude of other terms, see Fraser, 1999, for an extensive list of alternative terminology, are 'words or phrases that function within the linguistic system to establish relationships between topics or grammatical units in discourse' (Hellerman and Vergun, 2007: 158).

Discourse markers are seen to adopt a range of pragmatic functions within discourse, and operate in a similar way to backchannels, insofar as they help

to 'manage and negotiate topics' (Burns and Seidlhofer, 2002: 218), and are also used to show a mutual understanding, a 'shared knowledge' (Labov and Fanshel, 1977: 156), between the speakers. The fact that spoken discourse markers co-occur with both spoken and non-verbal backchannels is interesting as such phenomena may be seen to be functioning collaboratively across the speakers, helping to maintain their relationship and/or to jointly structure the discourse.

6.3.5.5. Summary

Again, the lines of enquiry undertaken in Section 6.3.5 have enabled some interesting observations to be drawn regarding the use of backchannels in discourse, based on evidence from real-life conversation. Specifically, this section has illustrated the manner in which head nods and spoken backchannels are often used together across turn boundaries, supporting the following statements:

R. Of the majority of the 1498 backchanneling head nods used with concurrent spoken backchannels in the corpus, *circa* 50%, 731, have a one-to-one temporal relationship between the vocalisation of the spoken form and the performance of the nod.

S. The remainder of the nods that co-occur with spoken backchannels, i.e. the ones without the one-to-one mapping of location, are most likely to be performed prior to the initial verbalisation of the concurrent spoken form. This is followed by nods that continue on from the verbalisation, and finally by nods that both precede and follow it.

T. There is no clear relationship between nod length and/or intensity and whether it is performed prior to/following the utterance of the concurrent spoken backchannel.

U. Spoken *and* non-verbal backchannels collocate with the use of grammatical discourse markers and deictic expressions in the corpus. While they are generally used more frequently with function than content words, no clear-cut pattern for such usage exists.

V. However, while listener backchanneling nods are often directly *followed* by the speaker's use of content words, i.e. nouns, verbs and adjectives, they are less frequently *preceded* by such lexemes.

6.4. Chapter summary

The analyses undertaken in this chapter have provided a worked example of how a particular linguistic phenomenon can be explored in MM data using a CL approach. The chapter has provided an in-depth examination of the characteristics of backchanneling behaviour, as witnessed in the five-hour MM corpus of dyadic conversational data. In response to the 10 premises set out in the introductory chapter, this chapter has shown that some clear patterns exist between the collaborative use of spoken and non-verbal backchanneling forms; listing 22 key observations as a means of mapping these patterns, from A to V. These results have enabled us to develop a detailed profile of the ways in which backchannels operate in discourse.

Although many of the findings presented within this chapter are not necessarily counter-intuitive, previous studies have failed to investigate to what extent these patterns hold true in real-life conversational contexts.

Previous studies either fail to incorporate a corpus-based approach when examining behaviours, conducting analyses across a large sample of authentic data, or they tend focus in detail only on either spoken or non-verbal behaviour, at times mentioning the other type of behaviour in passing. Other studies fail to adequately provide such an exhaustive account of the collaborative, simultaneous use of spoken and non-verbal backchannels, as the present study has accomplished.

In order to further discuss the relevance of the analyses undertaken in this chapter, Chapter 7 provides a detailed qualitative, discourse-analytical linguistic commentary on the relevance of the results and patterns seen.

Chapter 7: Examining the Findings

7.1. Introduction

This chapter provides a discussion of the relevance and importance of analyses undertaken in Chapter 6, outlining the extent to which the findings have contributed to the understanding of backchanneling phenomena. In short, the chapter will:

- Examine in more detail some of the most interesting findings sourced from the analyses, contextualising these comments using specific examples from the data.
- Discuss contextual and co-textual factors that may have contributed to specific patterns of results.
- Provide a linguistic coding matrix for defining and encoding spoken and non-verbal backchanneling forms in discourse, completing and extending the matrix presented in Figure 6.1 of Chapter 6.

7.2. Overview of findings

7.2.1. Backchanneling forms and functions

Chapter 6 revealed many interesting characteristics of the ways in which backchanneling phenomena are used in real-life discourse. The investigations began with a basic comparison of the forms and functions upheld by these behaviours, so as a starting point to this discussion it is relevant to briefly review related findings here.

Firstly, the results disclosed that the majority of spoken backchannels, as used by all speakers in the corpus, were of a highly conventionalised and standardised nature (supporting claims by Oreström, 1983; Tottie, 1991; Gardner, 1997b, 1998 and Rost, 2002, as outlined in Chapter 2). This is in terms of the basic lexical forms/structures, and in terms of the functions they were commonly used to fulfil. Despite the fact that 195 different structural varieties of spoken backchannels were found in the corpus (refer to Appendix 6.7 for more details); 24 simple, 12 double and 159 complex forms, it was only a small minority of lexical forms that were actually used at any real frequency by the participants. Thus of the 2340 spoken backchannels seen, the simple forms *yeah* and *mmm* were most prolific (see findings D, E and F in Chapter 6 for details), together accounting for 63% (1465) of the total.

As demonstrated with the 'top ten' most frequent forms in Figure 6.14 of Chapter 6, these simple forms, as with the common double forms, such as *mmm*, *yeah* and *mmm mmm* for example, were most often used as CON and CNV tokens. Whereas, complex and some double forms were often backchannels adopting ER or IR functions, those at the opposite end of the continuum of facilitative feedback (Stubbe, 1998b) to the simple forms (refer to findings G, H, I and J in Chapter 6). This pattern suggests a fundamental relationship between backchannel form and function, one that supports previous claims in the literature reviewed in Chapter 2.

Based on the results of the analyses, it can be suggested that non-verbal backchannels also behave in quite a similar, conventionalised manner in discourse. Nods can theoretically constitute a range of different movement structures; from simple single nods, to long combinatory nods, comprising

infinite sequences of intense and more moderate peaks and troughs. However, it was again a nod variety in its most simple form that proved most common in the corpus. Nods of short duration and/or with a low intensity, type **A** nods in particular, were most often used overall (see finding L for details). These were followed by low intensity nods of a slightly longer duration, that is type **B** nods, then short nods with a greater intensity, so type **C** nods.

The analyses also suggested that both spoken and non-verbal backchannels were utilised constantly throughout the course of a conversation, although the particular *rates* of use, and the ratio of use between speakers, was naturally highly variable, and to some extent idiosyncratic. So, while some participants may use more non-verbal forms, some may use more spoken forms; and others may use an almost equal amount of both (refer to findings A, B, C, D and K of Chapter 6 for details).

Nevertheless, supporting traditional conceptions of backchanneling phenomena, the results suggested that the participant who appeared to adopt a more 'passive' role in a conversation which is crudely defined in terms of the number of words used by a given speaker, was the one who generally used a larger number of backchannels than the other speaker (see findings B, C and D in Chapter 6 for details). This relates to the fact that backchanneling behaviour is inherently a listener activity.

These least vocal and/or least 'active' participants were also significantly more likely to use a higher amount of backchanneling nods than the other participant. This is true in terms of raw frequency; the net usage, and in terms of proportion of use, from the total number of spoken and non-verbal forms used by each speaker. This pattern of frequency was not *always* seen to be

the case with the spoken forms (refer to Figures 6.3 and 6.4 in Chapter 6 for details).

7.2.2. Backchannel use in time and co-text

In addition to the basic patterns in the frequency of use of backchanneling phenomena, Chapter 6 also revealed some interesting patterns involving the positions at which these behaviours were commonly located in talk.

Firstly, examinations of the lexical collocates of spoken and non-verbal backchannels (refer to Section 6.3.5.4 of Chapter 6 for further details) suggested that, when used in isolation, both spoken and non-verbal backchannels were often positioned close to, and indeed directly after, prepositions, pronouns and determiners in speech. Spoken forms were also frequently followed by such function words and rarely, at R1 in particular, by content words. Even the use of the simple backchannel *mmm,* prior to, or following, the use of a noun, verb or adverb for example, proved to be infrequent in this analysis (see findings U and V for details).

It can be assumed that the reason for this is that these function words provide a co-textual environment where a TRP can often be legitimately placed (refer to Sacks et al., 1974 and Cathcart et al., 2003). In other words, these may represent possible completion points of turns, positions where overlaps and backchannels commonly occur (see Sacks et al., 1974), as it is at these TRPs that the listener can move to take the floor without interrupting the conversational flow from the speaker (refer to the discussion in Section 2.3.1.3 of Chapter 2). In contrast, this crude distinction implies that content words, words with a more specific fixed lexical content, instead signal that the

speaker is still mid-turn, so the use of backchannels directly after such forms would provide more of a challenge to the co-operative nature of talk (Grice, 1989). Therefore, the collocation with function words is not wholly surprising. However, given that function words are more common, in general, in discourse it is difficult to fully qualify this claim.

Nevertheless, as a point of contrast it is interesting to note that the 'gesture phase' of backchanneling nods frequently commenced prior to use of content (as detailed in findings U and V), rather than function words in talk. So although function words are generally more frequent in discourse, and are commonly collocates of spoken backchanneling forms, this is not the case for the nods.

This finding supports the idea that backchanneling nods are perhaps more flexible in their positioning in talk than their spoken counterparts. Furthermore, it suggests that they are less threatening to the collaborative nature of talk (as discussed by Bublitz, 1988; Rost, 2002 and Allwood at el., 2007a, see Section 2.3.2.3 of Chapter 2 for more details). Regardless of whether these backchanneling nods are short, long, intensive or otherwise (see finding T for details), they provide less as a challenge to the turn and, thus, were frequently used within and across turn boundaries, at TRPs and beyond.

This finding also further strengthens the implication that the least frequent speaker, the one who adopts a more passive role in talk (refer back to 7.2.1 for details), is more likely to use a greater number of nods than their more vocal counterparts. The more dominant speaker, (s)he who holds the floor more frequently throughout, is less likely to provide TRPs for spoken forms to be performed. Thus, it is likely that nods are instead used to provide

backchanneling feedback, theoretically speaking, to maintain the flow of the conversation. In other words, the specific type of backchannel used by a listener corresponds directly with who has the floor at a given point in the conversation, and what position of the turn the talk is currently at. This implies that 'language and gesture take it in turns as to which one adopts the central role in a communicative event' (Norris, 2004: 2), depending on the particular characteristics of a given time and location in talk.

This pattern suggests that spoken and non-verbal backchannels are not strictly *interoperable* in discourse; that is, it is not necessarily the case that the use of these phenomena can be freely interchanged in talk. So, while a nod is quite flexible in terms of where it can be legitimately located, spoken forms are perhaps more fixed in their relative positioning.

7.2.3. Aligning the spoken and non-verbal

Despite this difference in location and co-textual position, the results suggested that spoken and non-verbal backchannels *do* fulfil similar semantic and pragmatic roles in talk. So although these backchannels are not strictly interoperable, evidence from the analyses suggests that they are highly *collaborative* and semantically synchronous.

Fundamentally, this is supported by the high frequency of spoken and non-verbal backchannel co-occurrence across the entire corpus, for every speaker and across each individual conversation (see finding M for details), thus suggesting a strong relationship between the manifestations of these behaviours.

This patterning of co-occurrence was seen across each structural form of spoken backchannels, that is, all forms, from simple to complex, were more likely to occur with than without nods. Nevertheless, this proved to be less consistent for the different functional categories of these spoken forms. So, while those backchannels functioning as CON and IR tokens, i.e. those at either end of the functional continuum, were consistently used more frequently with than without concurrent nods, the reverse was the case for forms functioning as CNV and ER tokens (refer to Figure 6.16 and finding O in Chapter 6 for further details). Although, in general across *all* spoken, as with non-verbal, backchannels, the rate of correspondence was significantly high.

In terms of patterning across spoken and non-verbal backchannels, it should be noted that the basic movement structure upheld by the majority of the backchanneling nods, which co-occurred with spoken forms, was in effect closely related to the basic lexical structure of such spoken forms. That is, longer and/or multiple nods (those with a relatively 'complex' movement structure, including types **B, D** and **E**, all of which were actually used with and without concurrent spoken forms at a fairly even rate, see finding P for further details) were most frequently used with double or complex forms of spoken backchannels rather than with simple single word spoken forms, as detailed in finding Q in Chapter 6.

Furthermore, the results suggest that in 50% of such instances, the location, i.e. the position in talk at which these concurrent backchanneling forms were used, the non-verbal forms were directly matched with the verbal counterpart (see findings R and S for details). So, with the backchanneling nod commencing and terminating at approximately the same time as the

verbal form, as seen in Figure 7.1 (using an excerpt from the transcript of supervision S01FM):

```
<$M> and I also I think that the so= conferences were definitely a success.
<$F> yeah
<$F> in just outright because I I enjoyed presenting and I +
<$F> yeah yeah
<$M> you know overall and er you know like you said I get some good feedback+
<$F> uh-huh
<$M> but also a success in terms of my own work.
<$F> yeah
<$M> in that it really gave me something to focus for on.
<$F> mmm
```

Figure 7.1: Concurrent spoken and non-verbal backchannel use- a basic one-to-one mapping.

This direct temporal mapping (refer to findings M and N for details) implies that at such points the spoken and non-verbal behaviours are adopting the same basic role, functioning in the same way, in talk.

In this example, there is a one-to-one relationship between the type of nod used (**C**, **C**, **B**, **C** and **A** respectively), and the lexical form and discursive function of the spoken backchannels with which they co-occur. In other words, the 'most minimal' types of spoken backchannel (O'Keeffe and Adolphs, 2008) i.e. *uh-huh* and *mmm*, both functioning as CON tokens, indeed co-occur with the 'most minimal' nod types (in terms of movement structure/ intensity), types **A** and **B**. Whereas, those forms functioning as more engaged CNV tokens, *yeah*, *yeah yeah* and *yeah* co-exist with a more engaged and emphatic nod structure, that is a type **C** nod. Although, such a pattern proved not to be strictly definitive in these results, as generally speaking all forms of

spoken backchannels, regardless of function, most commonly co-occurred with nods type **A** and **B**, due to the fact these were the most prolifically used types overall (see Figure 6.17 in Chapter 6 for details).

Nonetheless, this tendency for the mapping and co-occurrence of spoken and non-verbal forms adds an additional level complexity to the issue of how concurrent backchanneling phenomena across different forms and functions can be defined. Given these findings, an important methodological question primarily to be asked is whether, for example, the response *uh-huh* seen here, which conventionally functions as a CON, should still be classified a CON when there is a concurrent nod or not. This also prompts the question of whether the specific type of nod used affects the classification of such a structure. In other words, it questions whether the nod is simply complementing the concurrent form, or rather whether the addition of the nod changes the discursive properties of a conventional *uh-huh,* used in isolation. Does the concurrent nod reinforce and/or alter the pragmatic function and/or associated meaning of this spoken backchanneling response in some way?

A perceived change in the role of this spoken from, as part of a 'single collaborative backchanneling unit' therefore, suggests that the basic coding model, as offered by O'Keeffe and Adolphs (2008), requires revision. The question of what specific, different, function(s) such units are adopting instead cannot easily be answered, although this is further discussed in Section 7.3 below.

Beyond this one-to-one mapping, the analyses also revealed many instances where single backchanneling nods co-occurred across turn boundaries, thus with multiple spoken forms. Such instances effectively

provided a 'one-to-many' rate of co-occurrence. An example of this is seen in Figure 7.2, taken from the transcript of supervision S04MM (also see section 6.3.4 and Figure 6.10 of Chapter 6 for the related discussion).

```
<$2> +but no one would comment on+
<$1> Okay I've got you
<$2> +the choice between+
<$1> Yes,
<$2> +using it predicatively+
<$1> Yeah,
<$2> +or attributively+
<$1> Yeah right,
<$2> +and whether that the that would have any difference make any
difference+
```

Figure 7.2: Mapping spoken and non-verbal backchanneling functions.

When examining the transcript alone, it is possible to identify four instances of spoken backchannels here; *okay I've got you*, *yes*, *yeah* and *yeah right*, all of which can be assigned different discursive functions using the frameworks outlined in Chapter 2; ER, CNV, CON, and IR tokens, respectively. On replaying the video record of this episode, these four instances co-occurred with the same, single nod.

These backchannels are used sequentially over the speakers' turn, in quick succession over very short turns with only slight pausing between each (again, as evidenced by the playback of this episode using DRS). Given this, and the fact they co-occur with a single backchanneling nod, again questions whether these episodes should be considered as four single instances of backchanneling, or together as part of a larger, more global, MM backchanneling structure. Furthermore, it raises the question of which function, of any used, most appropriately defines the nature of this behaviour.

That is, whether it is best described as fulfilling a function of a CON, CNV, ER or IR token, or indeed none of these.

Questions of this nature also need to be raised for situations where the nod actually precedes and/or follows the point at which the concurrent spoken form is uttered. This phenomenon was again extensive in this corpus, and an example is provided in Figure 7.3; a transcript excerpt taken from S05MM:


```
<$1> +more grounded whereas a mobile phone's perhaps more casual and
<$2> Definitely yeah and well like I say just more expected more daily+
<$1> Mm.
<$2> +behaviour
<$1> Mm that's right. Yeah
```

Figure 7.3: Non-verbal backchannels preceding the use of concurrent spoken forms.

Here a type **B** nod is used with spoken backchannels adopting the CON and ER functions. Here the nod starts prior to the point where the first concurrent spoken forms, *mm*, is uttered, and then commences across turns before finally terminating at the same point where the second concurrent spoken phrase, *Mm that's right. Yeah* ends.

Overall, the results highlighted that the nods used across turn boundaries and multiple spoken forms, as seen in Figures 7.2 and 7.3, were consistently of types **B**, **D** and **E**, i.e. longer length nods. Of interest to note, however, is that the highest proportion of spoken forms functioning as ER and IR tokens used in such situations co-occurred most frequently with the more intense of these nods, i.e. types **D** and **E**. Whereas, type **B** nods were most frequently used with those spoken forms functioning as CON and CNV tokens (other

than the exceptions, as seen in Figure 7.2 and 7.3, where spoken backchannels adopting a multitude of functions were used with the nods). The only departures from this, albeit a minority across the corpus, are S02MM.1 and S03MF.M who both use an IR token with a type **B** nod across multiple backchanneling turns and S02MM.1, S04MM.2, S05MM.1 and S06FF.1 who use these nods with concurrent ER tokens. Having said this, finding T in Chapter 6, revealed that no definitive relationship emerged between the type of nod used and whether it was specifically administered prior to and/or following the utterance of the concurrent spoken form.

However, despite these exceptions, the general results again strengthen the suggestion that there is a basic relationship between the type of nod used with a concurrent spoken backchannel, and the basic discursive function adopted by this spoken form, not only within but across turn boundaries. Therefore, the dynamic, highly variable and to some extent idiosyncratic nature of both nods and spoken backchannels suggests that 'rather than using discrete categories such as convergence or continuer function it may be more appropriate to conceive of backchannel functions as a cline that moves from a simple continuer function to an engaged function as one of the possible axes' (Knight and Adolphs, 2008).

In short, this supports the notion that there is a certain level of synthesis in the pragmatic functions of backchannels across the different modes of communication. So although, as identified in section 7.2.1, a specific motion of the head is not *exclusively* fixed to a specific word every time it is spoken, paradoxically, these spoken and non-verbal backchanneling forms still have the capacity to operate simultaneously, that is, mutually, in conversation.

As an extension to this it is of interest to note that, in the example given in Figure 7.2, the final spoken backchannel actually operates as an IR token. By definition, these tokens function to signal some sort of closure in the conversation, where information has been received and the listener prompts, for example, a topic shift or change. Therefore, given that it is used here at the approximate point that the concurrent backchanneling nod ceases suggests that the nod is operating in parallel with the spoken backchanneling token. The end of the movement of the extended nod may also signal a point of closure; where information has been successfully received. To this extent the nod perhaps adds emphasis to the associated pragmatic meaning of the concurrent spoken word(s).

However, while there are many other instances where this pattern is seen across the corpus, as documented in Appendices 6.11 and 6.12, this is not true in every case. On other occasions, the nod movement follows far beyond the exact point at which, for example, the IR token is used. In general, no definitive relationship between whether the nod preceded and/or followed concurrent spoken forms across turns, and so on, was found.

In relation to this notion of the gesture phase, it is appropriate to mention that the study has perhaps neglected to examine, on a more finite scale, the exact point, for example, that the *most emphatic* or pronounced part of a nod is performed throughout a stretch of talk (refer to section 2.4.2.2 of Chapter 2 for discussions related to the gesture phase, also see Kendon, 1982). This information may have helped to answer many of the questions concerning the relationship between spoken and non-verbal use that have arisen in this

chapter, however, the notion of the nod 'stroke' was not an immediate concern for this study.

This is because, again, when analysing gesture (refer back to Section 2.4 of Chapter 2 for related discussion), especially when using those manually ascribed techniques used here, it is practically impossible to achieve 100% accuracy in defining and appropriately categorising isolated elements of a head nod movement, especially without multiple passes by a range of different 'raters' (see section 4.2.3.1 of Chapter 4). Indeed the challenge of specifying, for example, the start and end of a nod's movement phase, as explored above, is sufficiently complex without attempting to explicitly define the most emphatic point in this phase.

Even short and/or less intense nods, including types **A** and **C**, can have a 'fuzzy' phase structure making their initial detection and classification heavily subjective, thus open to scrutiny. Types **B**, **D** and **E** nods are obviously even more problematic to tackle as their phrase is even less discrete and freely identifiable given their complex structure and the fact they potentially comprise a range of different, multiple episodes of sequentially structured individual nods. In these cases, it is debatable whether such clusters of nods should instead be separated into smaller movements, with individual head-up peaks being classified as the 'most empathic' of each of separate nod segment.

The problem with using such an approach is there is no guaranteed cost-benefit, insofar as it cannot be assured that this process will provide results that are any more informative or accurate than those already gained.

Therefore, it cannot be wholly guaranteed that this is necessarily a better alternative, and a solution to the problems posed.

In sum, by 'aligning the spoken and non-verbal', in the analyses, a range of theoretical and methodological questions regarding defining and classifying backchanneling behaviours have been raised; questions that challenge and extend existing linguistic models for classifying such phenomena. These include, for example, whether a single nod across multiple turns and/or within turn boundaries should be classified as fulfilling a single function in the discourse; whether concurrent, multiple spoken forms of such nods should be regarded as adopting a function that is analogous to each other and/or the concurrent nod, rather than adopting a range of different functions, and so on. These questions possibly foreground one of even more importance, that is, 'what is a backchannel?' which will be readdressed in the next section, Section 7.3.

7.3. A coding matrix for multi-modal backchanneling phenomena

7.3.1. Introducing the matrix

In order to fully illustrate what this study has added to the knowledge of backchanneling behaviour, as detailed in previous research, an adapted pragmatic-functional coding matrix of these phenomena is offered here. This assists in conceptualising elements of the non-verbal in conversation in relation to, and synchronicity with, spoken features.

Theoretically speaking, it is difficult to fuse the different, dynamic properties of these phenomena in a single, integrated conceptual matrix. The relative success of this process is determined by, for example, questions of

whether it is indeed appropriate to describe the spoken and non-verbal elements as single backchanneling units, or whether to initially frame behaviours across the modes in a distinct way, indicating correspondences and similarities across these where relevant.

Furthermore, it is problematic to determine which, if any, specific behavioural characteristic(s) are perhaps more important than others to this conceptualisation, and so to assess which elements of these behaviours are more important for consideration than others during classification. This essentially questions, for example, whether the location of, a head nod, its form, or the type of backchannel with which it co-occurs (if relevant), is more significant in this classification. Since it is difficult to provide a definitive answer to this question, it is premised that instead it is more appropriate to observe, as with the study in Chapter 6, a combination of these factors, mapping one to another, as a means of classification.

To date, the thesis has essentially been concerned with investigating five key elements, properties of backchanneling behaviour, as listed below (building on the matrix offered in Figure 6.1 of Chapter 6). These have proved to be most decisive in helping to define 'what is a backchannel?' in discourse. The principal findings associated with each of these properties, as supported by the MM corpus analysis, are also summarised under each.

- **Frequency**: Rate of occurrence.

  - All backchannels, across the modalities, are frequent and used at a constant rate in interaction.

  - The specific rate of use by a given participant depends on their role at that point in the conversation. The rate of use is at its highest with a 'passive' listener, and its lowest with the more 'active' speaker.

- **Location**: In terms of immediate lexical and behaviour co-text, observed through collocation searches and scatter plots.

  - Nonchalant nods and/or nods of a short duration are flexible in their positioning in discourse. Spoken forms, both those used in isolation, and with concurrent nods, are more closely tied to TRPs in their positioning.

- **Form/structure**: The basic lexical form and/or movement shape, from short to long, simple to complex (i.e. lexis and nod type).

  - Both spoken and non-verbal backchannels, when used in isolation, or indeed in conjunction, are most prevalently short in form, so of a simple lexical structure, or of a short duration (as with nods types A and C).

  - Individuals have a tendency to use and re-use the same simple form(s) throughout the course of a conversation, although the specific lexeme used is subject to variation from each individual to the next, so while, for example, one speaker may use the simple form *yeah* most often, another may use *mmm*.

- **Concurrent behaviours:** Observing the wider context of the behaviour, i.e. the type(s) of lexical content and/or nods that co-exist with specific instances of phenomena.
  - Spoken and non-verbal backchannels are highly collaborative, frequently co-existing throughout the course of the conversation.
  - The location mapping of these concurrent behaviours is most often of a one-to-one nature, although it is not always restricted to this, as many instances of one-to-many nods to spoken forms exist.
- **Function**: The interactive, pragmatic function of the backchannel-in-use, that is, the task(s) which it performs in each given instance.
  - In general, nods used alone (especially those which are short and/or nonchalant) are the most minimal forms of backchannel, posing the least threat to the flow of conversation. These nods are, therefore, closely aligned with spoken forms adopting the CON function.
  - Backchanneling nods used concurrently with spoken forms, matching their location in a one-to-one nature, assume the same discursive function as this spoken form, although these instances are possibly more emphatic than instances of single spoken and non-verbal forms. These nods are perhaps most appropriately aligned with the CNV function.
  - The pragmatic function of nods across turn boundaries and multiple concurrent spoken forms is more variable, and reliant on the co-text and context of use.

At this point, it is important to note that the last of these elements, the pragmatic function of backchannels, is the category which has proved to be the most problematic to tackle to date, both in the previous literature and in the study analysis.

Overall, while this study has effectively highlighted that it is difficult to directly match kinesic forms of non-verbal backchannels with a particular discursive function (based on current coding schemes, insofar as they lack the utility for such definition), it is shown that there is perhaps a close relationship between the function of a nod movement and its use in relation to spoken forms. That is, whether or not it co-occurs with spoken backchannels, and on the particular form etc. of this lexical unit. Although these patterns are not necessarily counter-intuitive, it was not possible to support such claims when using traditional mono-modal corpora. It is only in MM corpora with the integration of video that these patterns can be fully supported. Therefore, while the previous literature has, in passing, made reference to such patterns (see Section 2.4.2.3 of Chapter 2 for more details, also see for example Maynard, 1987; McClave, 2000 and Norris, 2004), these have never been extensively investigated in the way that the current study has done.

This finding effectively completes the '**?**' section of the coding matrix seen in Figure 6.1 of Chapter 6. Again, these functions exist in the form of a cline, in effect from the most minimal to the more engaged forms of non-verbal backchannels. While, for example, nods used in isolation are general the most minimal, least imposing, forms of backchannels, this is perhaps more true for nods which are low in intensity. More intense forms may act more emphatically, so function in a more engaged way insofar as they are likely to

be more noticeable to the speaker, and may act as providing feedback (in a comparable way to CNV tokens), rather than merely maintaining the flow of talk.

Naturally, there are many exceptions to these basic patterns and, as discussed briefly in Section 7.3.3, there are naturally many shortcomings associated with any attempt to model forms of gesture in natural conversation.

To summarise this section, Figure 7.4, overleaf, presents these properties in a theoretical matrix, a coding scheme, which can be used when defining and examining spoken and non-verbal backchannels in real-life discourse.

While the figure is structured, as with the O'Keeffe and Adolphs model (2008) according to four key pragmatic functional categories (refer to section 2.3.2.2 of Chapter 2 for definitions of each of these functions), this is simply for ease of reference. These categories are by no means taxonomic as a certain amount of overlap and inter-changeability is possible within and across these boundaries. Again, the content of this matrix is effectively hierarchically structured, with different elements being ordered according to the complexity of their structural form(s) and associated semantic content, and the frequency with which they are often used in discourse. In other words, positioned at the top of the continuum are the most frequently used forms of backchannels seen in the corpus; the low intensity, short duration nods, and the simple form backchannels with minimal lexical content and relational value (refer to Chapter 2 for further discussion).

Pragmatic function	Structural form(s) and common examples	
	Backchanneling Nods	**Spoken Backchannels**
CON	Nods of a low intensity and short duration and nods of a low intensity and long duration, used within speaker turn boundaries. Examples: Type **A** and **B** nods	Predominately simple, but some double lexical forms of spoken backchannels (i.e. derivations of the most common simple forms). Examples: *mmm, yeah, mmm mmm, mhm, uh hm*
	One-to-one concurrent simple form spoken forms with low intensity nods of a short duration, used at TRPs. Nods map the relative start and end of the spoken form. One-to-one concurrent simple form spoken backchannels, and nods of a low intensity and long duration, used across turn boundaries. Nods tend to directly map, in terms of location, the start and end of the spoken form. Examples: Type **A** nod + *mmm*	Type **A** nod + *yeah*	Type **B** nod + *mmm*	
CNV	Nods of a high intensity and short duration and some nods of a low intensity and long duration. Examples: Type **A**, **B** and **C** nods	Simple and double lexical forms. Examples: *yeah, yeah yeah, yes, yeah okay*
	One-to-many concurrent simple or double form spoken backchannels used with nods of a low intensity and long duration, or high intensity and short duration. These nods are used across turn boundaries. In such instances all spoken backchannel forms are functioning as CON and/or CNV tokens. Nods may precede and/or follow the use of the concurrent spoken form. Examples: Type **A** + yeah (CNV)	Type **A** + *yeah* (CNV), *mmm* (CON)	Type **B** + *yeah yeah* (CNV), *yeah yeah* (CNV)	Type **C** + *yes* (CNV)*, yeah* (CNV)	
ER	Nods of a low intensity and long duration, or of a high intensity and short duration. Examples: Type **B**, **C** and **D** nods	Double and complex lexical form backchannels. Examples: *definitely, yeah absolutely, that's right, yeah that's right*
	One-to-many concurrent simple, double and/or complex form spoken backchannels used with nods of a low or high intensity and long duration across and/or within turn boundaries. Each spoken backchannel is either functioning as ER tokens, or a combination of ER, CNV and/or CON tokens. Nods often precede and/or follow the use of the concurrent spoken forms. Examples: Type **B** + *mm* (CON) and *mm that's right. Yeah.* (ER) (see Figure 7.3)	Type **C** + definitely (ER)	Type **D** + *yeah absolutely*	
IR	Combinations of nods with a long duration, high intensity and/or low intensity, or nods with a short duration and high intensity. Examples: Type **C**, **D** and **E** nods	Some common simple form and complex form backchannels. Examples: *right, okay, right okay, yeah okay, sure*
	One-to-many concurrent simple, double and complex form spoken backchannels (a combination of these), used with nods of either a high intensity and long duration, or a long duration with a combination of nods of a high and low intensity. These nods are used across and/or within turn boundaries. In such cases each concurrent spoken backchannel is functioning as an IR token, or a combination of IR, ER, CNV and/or CON tokens. Nods often precede and/or follow the use of the concurrent spoken forms. Examples: Type **E** nod + *okay I've got you* (ER), *yes* (CNV), *yeah* (CNV), *yeah right* (see Figure 7.2).	

Figure 7.4: A coding matrix for examining the relationships between spoken

and non-verbal backchannels in discourse.

306

Following this are concurrent forms of these phenomena. Although, overall, such concurrent simple spoken and non-verbal forms were most common in the corpus, the use of these in unison may be perceived as providing feedback that is more emphatic than, for example, a nod is isolation thus are positioned accordingly lower in the matrix.

Positioned at the bottom of this continuum are again more engaged forms of backchannels. These forms are least frequently used in discourse and comprise complex structures; ranging from simple, single lexical items used in isolation to combinations of low and high intensity nods of a long duration used within and across a multitude of spoken forms, each either adopting the same ER function, or a range of different discourse functions.

7.3.2. Limitations of the matrix

Although this coding matrix is directly supported by the corpus data examined in this thesis, that is, it is transferable across each of the speakers in each of the conversational contexts, as discussed in specific instances above (see Section 7.2), this is not always necessarily the case for every conversational episode. Variations, i.e. anomalies in the basic properties upheld by specific backchannels, as outlined in the coding matrix, are possible at various levels, from the personal, i.e. idiosyncratic, to wider, discourse-contextual levels. For instance, gestures similar to words, may in fact 'be tailored for a particular addressee, in a particular conversation' (Bavelas, 1994: 206), in particular socio-cultural contexts, beyond the dyadic academic supervisory meeting environment observed here. Therefore, caution must be exercised before applying this coding scheme broadly to other MM datasets.

As with any coding and/or behavioural classification model, this matrix somewhat 'obscures complexity and differences [of real-life interaction], leading to generalisations that are insensitive to subtle differences dependant on sequential position in the floor of the interaction and individual speaker differences' (Gardner, 2001: 17). Again, this is because real-life discourse is spontaneous and to some extent transient, therefore, it can never be possible to fully conceptualise these widely various behaviours. Thus, these can never truly be delineated as part of a theoretical conception. Instead, these matrices/models, at best, offer insight into some of the ways in which these phenomena behave in discourse, providing a useful method for increasing understanding of how such elements are used to generate meaning in talk.

Other more general limitations of the actual study, and the results derived from this, are discussed in Chapter 8.

7.4. Summary

This chapter has re-examined the question 'what is a backchannel?', detailing how the definition of spoken and non-verbal forms of this phenomena has changed or been enhanced in light of the results gained from a MM analysis of these behaviours.

Through the examination of these MM facets of backchanneling phenomena, this study has not only extended the current understanding of these, but also actively questioned whether it is possible, within a CL methodological framework, to describe and analyse characteristics of language and gesture use together. Given the discussions undertaken both in this chapter, and Chapter 6, it is suggested that language and gesture-in-use

can be analysed and described together, and it is with the change in modality, that is, the addition of the non-verbal perspective in the MM corpus which has enabled this. However, whether this finding is restricted to instances of backchanneling phenomena alone remains to be seen.

The following chapter provides the conclusion of the thesis, giving an overview of the principal concerns and findings from Chapters 1 to 7, and presents a critical review of MM corpora, and the MM CL approach that has been used in the main study. Based on these discussions, the chapter furthermore offers suggestions for the ways in which this approach can be further adapted and refined for use.

Chapter 8: Conclusion

8.1. Thesis overview

This thesis began with the aim of investigating the following:

1) *Developing the next generation of linguistic corpora*: What are the key technical, practical and ethical issues and challenges faced in the design and construction (i.e. the 'development') of MM corpora, and how can these best be approached?

2) *Using MM Corpora*: What are the roles, forms and functions of non-verbal and spoken backchanneling behaviour in real-life, naturally occurring discourse, and what is the relationship between them?

In order to provide a background for these, Chapter 2 presented a detailed argument for the potential strengths of emergent MM corpus datasets for linguistic enquiry and, working from a bottom-up perspective (for addressing the first concern), Chapter 3 complemented these discussions by providing a detailed account of the *how* of designing and constructing MM corpora.

It explored the principal technical, practical and methodological issues associated with recording, coding and (re)presenting MM records, discussions which were augmented with a more generalised commentary of the *functionality* of MM corpora; an identification of the significant challenges faced when attempting to make MM datasets 'usable' for the corpus-based researcher. This notion of functionality was the key concern of Chapters 4 and 5, which provided an extensive analysis of a ten-minute case study

extract in order to demonstrate the potential of the MM CL approach for investigating facets of the relationship between language and gesture-in-use.

Chapters 6 and 7 then presented an extended study examining the ways in which spoken and non-verbal backchanneling behaviours (co)exist and co-operate in talk. The analysis of this data revealed many interesting facets of the relationship between, and to some extent the co-dependency of, spoken and non-verbal backchannels in discourse. This assisted in complementing, and, importantly, extending what is already known about the use and function of these phenomena. A coding matrix to assist in the examination of MM backchannels is offered in Chapter 7.

8.2. Framing the findings

8.2.1. Corpus pragmatics

This thesis has operated on the notion that in order to fully assess the importance, i.e. the 'added value' of the findings from Chapter 6, it is necessary to complement these descriptive results; which have primarily focused on classifying the 'meaning of the actual language form or 'sign' used', with 'other sources of knowledge, such as knowledge about the context of the situation, knowledge about other speakers or listeners and knowledge about culturally recognised norms and activities' (Knight and Adolphs, 2008).

Therefore, it attempted to integrate these principles of pragmatics, i.e. 'the science of the relation of signs to their interpreters' (Morris, 1946: 287), patterns of *meaning*, with the quantification and interpretation of patterns of actual language–in-use across large scale datasets, that is, supported by CL methodologies.

When using text corpora, it has been difficult, practically impossible, to fully combine an investigation of patterns of discrete items, units of behaviour (as is common to corpus-based analysis), with such descriptive frameworks of the functions of these units (as is common to pragmatics) when analysing real-life interaction (see Adolphs, 2008: 6-8). This is because contextual information is effectively 'missing' (see Widdowson, 2000; Mishan, 2004 and Mautner, 2007) in such 'units' of behaviour when using a basic CL approach because discourse is stripped to its lowest common denominator in corpora; i.e. that of text.

While metadata and other forms of data coding can help to inform of the identity and biographical profile of who was speaking to who and where this conversation took place, this supplementary record effectively still 'presents no more than a trace of the original discourse situation' to which it pertains (Thoutenhouft, 2007: 3, also see Adolphs, 2008 and Knight and Adolphs, 2008). This is because, as discussed in section 3.4 of Chapter 3, the reality of the discourse situation is lost in its representation as text, so for example communicative gestures and paralinguistic properties of the talk are lost during this process.

This limitation is problematic because an understanding of the context of interaction is not only vital for the investigation of vocal signals, the words spoken (see Malinowski, 1923; Firth, 1957 and Halliday, 1978), but also for understanding the sequences of gesture: 'just as words are spoken in context and mean something only in relation to what is going on before and after, so do non-verbal symbols mean something only in relation to a context' (Myers and Myers, 1973: 208).

Therefore, an advantage of the current study, and indeed of MM corpora in general, is that the integration of the video and audio data allows for an additional representative dimension of the reality of the context (refer to Section 2.2.4 of Chapter 2 for further discussion related to discourse context). Fundamentally, the video provides a more lifelike representation of the individual and social identity of a participant, allowing for an examination of prosodic, gestural and proxemic features of the talk in a specific time and place. It reinstates partial elements of the reality of discourse, giving each speaker and each conversational episode a specific distinguishable *identity*. Thus, in short, the thesis has operated on the premise that it is only when these extra-linguistic and/or paralinguistic elements are represented in records of interaction that a greater understanding of discourse beyond the text can be generated.

However, even such records are arguably partial in their representation of context, because context combines not only 'extrinsic'; 'social, cultural and interactive' factors, but also include 'intrinsic'; 'cognitive, affective and conative' factors (Kopytko, 2003: 45, also see Labov, 1972; van Dijk, 1977; Duranti and Goodwin, 1992; Eckert and Rickford, 2001 and Fetzer, 2004 for further discussion on language and context). In effect, 'the scope of interactional context is indefinite and infinite because each context is embedded in its own context that is embedded in its own context and so on', creating a theoretical 'situation of infinite contextual regress' (Kopytko, 2003: 50). This suggests that it is impossible to fully capture this notion of 'context' as the abstract and indefinite definition of context does not actually lend itself to this.

So, while the addition of video in the MM corpus can arguably allow for a richer description of some of the extrinsic contextual features of interaction, it is difficult to fully quantify, qualify and interrogate *all* such features in a meaningful way. This problem is intensified with the intrinsic features of context insofar as these are not freely observable, even when utilising MM of the nature described in this thesis. This exists as a current, principal theoretical and methodological challenge for corpus pragmatics and context.

Therefore the sections in chapter 7, in particular, did not strictly seek to make impressive generalisations, or propose definitive schema for defining and examining backchanneling behaviour in discourse. It is not intended to provide a prescriptive grammar of backchannel use across speakers and contexts, as this aim is perhaps impossible to achieve and misguided in its focus. This is especially true in light of the potential partiality of the 'reality' of real-life discourse, provided by the physical corpus and associated CL methods used. Instead, Chapter 7 provided a prospective account of the ways in which backchannels *appear* to function and operate in the specific discursive episodes seen in the corpus, in order to complement and extend what is already known about this discursive phenomenon.

8.2.2. Language, gesture and cognition

The study emphasised that, although many general patterns of backchanneling behaviour were found in the analyses, it should be recognised that in many instances the specific facets of this phenomena, as adopted by individual speakers was, to some extent, idiosyncratic and/or

highly variable depending on a range of individual, co-textual or, again, wider contextual factors.

It is difficult to determine to what extent these individual factors are more likely to influence, when the simple spoken form *yeah* is used, for example, and whether it is accompanied by a nod, at a given point in time. Similarly, it is problematic to determine the specific intentions of the listener, and whether they are even, themselves, consciously aware of their behaviour and the potential pragmatic meaning of using these backchanneling structures and/or combinations of these, in talk. Although it is possible to discuss patterns of the co-occurrence of backchannels behaviours, at present very little is known about the cognitive processes behind these. This again relates to the more intrinsic aspects of the discursive context, the 'cognitive, affective and conative' elements that are inherent to natural human interaction (Kopytko, 2003: 45).

McNeill articulates the need for the 'full cognitive representation' (1992: 254) when describing gesticulation in use, and this is something which is lacking here. Generally speaking, this is something that is deficient across the landscape of conventional CL research, because CL based methodologies concentrate primarily on examining patterns in records of discourse and behaviours as they have been *produced,* i.e. they examine the *results* of linguistic production. They operate on the premise that 'gestures mainly serve an external function in communication' (Lozano and Tversky, 2006: 47).

Thus, corpora and CL methodologies allow the analyst to observe either how gestures facilitate the expression of the message, complementing or sometimes even opposing the spoken content of the message, or how

gestures are received by the speaker and/or listener. They do not lend themselves to the examination of the processes behind these results, nor do they assist the examination of how these behaviours are received, either consciously or subconsciously, and subsequently attended to, if at all, by the speaker in order to facilitate structured collaborative talk. Given that these are relatively abstract factors, they are difficult to capture and quantify in text and/or even in MM records of communication as offered in CL methodology.

Beyond the scope of CL methodology, however, there is an abundance of research on the fundamental relationship between language, gesture and cognition. Some of this research seeks to examine whether gestures function primarily to facilitate the retrieval of certain lexical forms in the mental lexicon of a speaker (known as the Lexical Retrieval Hypothesis, see McNeill, 1992), with the belief that 'gesture, together with language, *helps constitute thought'* (McNeill, 1992: 245, also see Krauss et al., 1991 and Krauss et al., 1995). Other studies instead explore whether gestures are used by a speaker to aid the listeners comprehension of a message (know as the Information Packaging Hypothesis, Alibali et al., 2000, 2001), or whether it is a combination of these and/or other factors (Kendon, 1994 and Alibali et al., 2000).

The approaches used in order to conduct studies of this nature, and indeed some of the results sourced directly from these studies, can be appropriately adapted and utilised to inform and extend current conventions in CL methodology. This would allow for a more cognitive investigation of the processes behind patterns of language and gesture-in-use.

However, in order to fully support this line of research, the approaches to recording, coding, (re)presenting natural language data again obviously require a complete redefinition. These concerns are beyond the scope of the current study, although such a 'symbiosis of cognitive linguistics and CL' has been cited as a priority for MM CL methodology (Gries and Stefanowitsch, 2007).

8.2.3. Limitations of the study

Beyond providing a cognitive perspective of backchanneling phenomena, there are a variety of other ways in which this thesis can be extended.

Firstly, as an early study of its kind, there are various areas of interest that are beyond the remit of the specific focus of the main study. Therefore, extensions to the study described in Chapters 5 and 6 could include:

- A more detailed examination of the complexities of the relationship between various other backchanneling 'cues at different linguistic levels, such as prosodic units, pitch contours, morphological categories, discourse markers or gaze direction' (Bertrand et al., 2007: 1).
- An investigation of the ways in which other forms of gesticulation interact with spoken language and/or other non-verbal phenomena in order to generate meaning (such as iconic hand gestures and their relationship to discourse markers, see Knight et al., 2009 for preliminary discussions of this specific enquiry).
- Examining the relationship between spoken and non-verbal backchanneling and pause phenomena.

- Explorations of idiolect differences in the use of spoken and non-verbal forms of backchannels, or indeed between other aspects of language and gesture-in-use.

Furthermore, in order to make this study manageable, it was necessary to be selective, with regard to the amount of data analysed, the conditions under which this data was collected, who/ how many participants were included, and so on (i.e. elements of the corpus design, as detailed in section 3.3 of Chapter 3). Thus, the focus of this study was restricted to a relatively small data set in terms of number of words spoken, number of participants involved, the range of discourse contexts examined and the age and status of the speakers. It only included conversations with British native speakers in a pedagogic background, making it difficult to determine to what extent the findings are transferable beyond this context. Therefore, an examination of a wider range of speakers from different socio-cultural and discourse contexts (from dyadic to group conversation) would also extend the focus of the current study.

However, in terms of practicality, and the time allowed for a PhD study it would be difficult to extend the focus in order for it to fully investigate each and every one of these areas, (and indeed the advantages of doing this are, in terms of cost-benefit, questionable). This is because the heavily manual approach that has been used for detecting, quantifying and encoding forms of backchannel behaviour only really supports the study of relatively small size datasets and/or sub-sets of behaviour of interest.

Given this limitation of the manual approach to analysis, in Chapter 5 it was recommended that, in order to enhance the potential of MM corpus methodologies, it would be useful to automate the process of MM CL enquiry. This is to reduce the amount of time required by the analyst when trawling through the video and textual data, and, thus, to increase the accuracy of any analyses. This chapter tested a digital tracking device which was designed to allow users to automatically define and subsequently encode specific features of interest in video data (according to specific parameters set by the analyst). In theory, this tracker should allow for larger scale explorations of language and gesture-in-use to be undertaken with ease (including studies whose focus is beyond backchannels and head nods), although in practice, since such technologies are still 'developing', these tracking techniques are far from perfect, so at present remain a speculative *potential* rather than *functional* part of the MM CL approach.

Apart from the tracker, the thesis has underlined a range of other features of the actual composition of early MM corpora that could also benefit from future redevelopment (i.e. what 'data' goes in to a MM corpus). As identified in Chapters 2 and 3, one of the main criticisms of the NMMC and many of its contemporaries (refer to Figure 2.1 in Chapter 2 for examples of other MM corpora), is the fact that it is relatively small in size, particularly in comparison to multi-million-word mono-modal corpora that are available. Further, it includes data extracted from only a small number of speakers, again in a specific discursive context, from a particular vantage point (i.e. according to the static positioning of the camera and/or microphone).

The natural next development in construction, therefore, relates to enhancing the variet*y* of data included in MM corpora, to enable the linguist to make better informed observations of language-in-use from a multitude of different perspectives. This needs to be complemented by better metadata descriptions and systems for integrating and annotating data across the modalities. Despite the advances made in this study, and similar ones of this nature, Blache et al. note that (2008: 1):

> We still need a linguistic theory taking into account all the different aspects of multimodality, explaining in particular how the different linguistics domains interact, at the same time we need to specify a standardised way of representing multi-modal information in order to give access to large multi-modal corpora, as richly annotated as possible.

As identified in Chapter 3, the DReSS II project aims to lay the foundations for such enquiries to be undertaken. This project seeks to allow for the collection and collation of a wider range of heterogeneous datasets for linguistic research, in order to enable the construction of richer descriptions of language use in relation to context. To achieve this, the project is focusing on collecting 'data' records of a range of everyday (inter)actions, including SMS messages, MMS messages, interaction in virtual environments (instant messaging, entries on personal notice boards etc), GPS data, face-to-face situated discourse, phone calls and video calls.

The compilation of such heterogeneous data may enable us to extrapolate further information about communication across a range of different speakers, mediums and environments. In theory, this could assist in the questioning of the extent to which language choices are determined by different spatial, temporal and social contexts in communication.

However, in reality, there are obviously a whole host of ethical, practical and methodological problems that need to be faced when constructing such corpora. Indeed, such problems may deter linguists from attempting to create MM corpora of this nature because, to date, simple solutions to these problems have failed to emerge. This includes matters of what and how behaviours are quantified, queried and represented to the linguist, and how patterns are statistically assessed and/or analysed; thus, developing the key areas of discussion addressed in Chapter 3.

The realisation of the these aims for heterogeneous multi-context corpora (or indeed even improved 3rd generation MM corpora of the nature discussed in this thesis) are heavily reliant on technological advancements; on the constant refinement of systems that will enable the capture and structuring of natural language-in-use, as well as software that will promote the interrogation of different MM datasets. Constraints attributed to questions of scalability are obviously inherent to the practical implementation of this 'next-step', since, as identified in Chapter 4, the processes of recording, transcribing, time-stamping and coding data remain very time-consuming, even with the utility of software such as DRS (and with the implementation of automated methods, as addressed in Chapter 5).

These prospective technological advancements are, in turn, reliant on institutional, national and international collaborative interdisciplinary and multidisciplinary research strategies and funding, because 'modern research is increasingly complex and demands an ever widening range of skills…..often, no single individual will possess all the knowledge, skills and techniques required' (for discussion on the advantages of cross and multi-disciplinary research see Newell, 1984; Katz and Martin, 1997 and Golde and Gallagher, 1999: 281).

8.3. Summary

In consequence of the research presented in this thesis, it is suggested that although 'the reality of language is [perhaps] too complex to be described *completely*' (Chomsky, 1957: 16), with the onset of *developing* MM corpora and MM CL methodologies, the means with which we are able to fill in at least some of the gaps in our understanding of the complexities of human discourse are now being presented. The integration of multiple modes of information, as offered by MM corpora, is instrumental in this advancement by providing a more multi-dimensional landscape for exploring elements of real-life conversational data (across a range of different *modes* of representation), beyond that offered by mono-modal corpora.

Guide to CD-ROM Data Disk

File 4.1   A video example of type **A** backchanneling nods (from S03MF.M).

File 4.2   A video example of type **B** backchanneling nods (from S03MF.M).

File 4.3   A video example of type **C** backchanneling nods (from S03MF.M).

File 4.4   A video example of type **D** backchanneling nods (from S02MM.2).

File 4.5   A video example of type **E** backchanneling nods (from S02MM.2).

File 4.6   Complete case study transcript (taken from the NMCC).

File 5.1   Frame-by-frame tracking outputs for case study speaker S03MF.M (with y-axis position and head angle).

File 5.2   Real-time head tracking output of the case study data, for S03MF.M (aligned with the original input video).

File 5.3   Frame-by-frame tracking outputs for case study speaker S03MF.F (with y-axis position and angle of the head).

File 5.4   Real-time head tracking output of the case study data, for S03MF.F (aligned with the original input video).

File 6.1   Annotated transcript for S01FM.

File 6.2   Annotated transcript for S02MM.

File 6.3   Annotated transcript for S03MF.

File 6.4   Annotated transcript for S04MM.

File 6.5   Annotated transcript for S05MM.

File 6.6   Annotated transcript for S06FF.

File 6.7   Key lexical collocates of backchanneling head nods across the five-hour corpus.

File 6.8   Examining the lexical environment of backchanneling head nods in the five-hour corpus: concordance outputs using Wordsmith Tools.

File 6.9   Examining the lexical environment of backchanneling head nods-charting lexical items frequently used in the discursive environment of nods.

File 6.10   Key lexical collocates of spoken backchannels without concurrent nods across the five-hour corpus.

File 6.11   Examining the lexical environment of spoken backchannels without concurrent nods: concordance outputs using Wordsmith Tools.

File 6.12   Lexical items frequently used in the close environment of spoken backchannels without concurrent nods.

File 6.13   Key lexical collocates of spoken backchannels (with concurrent nods) across the five-hour corpus.

File 6.14   Examining the lexical environment of spoken backchannels with concurrent nods: Complete concordance outputs using Wordsmith Tools.

File 6.15   Lexical items frequently used in the close lexical environment of spoken backchannels with concurrent nods.

Appendix 4.1: A breakdown of the forms and functions of spoken backchannels in the case study data.

KEY:
Note- this key is valid for appendices 4.1 to 4.6 and 6.1 to 6.18

DATA DETAILS:	KEY:
DATA DETAILS: **Male Supervisor =** <$M> **Female Supervisee =** <$F> **10 minutes, 2156 words (other characters removed for count)** **14.40 – 15.40 on supervision video (S03FM)**	**KEY:** **Spoken Backchannels (BCs):** <mark style="background:yellow">Continuers</mark> (CON) <mark style="background:cyan">Convergence Tokens</mark> (CNV) <mark style="background:magenta">Engaged Response Tokens</mark> (ER) <mark style="background:red">Information Receipt Tokens</mark> (IR) **Head Nods:** ___ - Nod with speech: <$M> **X** – Nod without speech: <$M> ___ - Nod with speech: <$F> ___ - Joint Nods, co-occurring with speech: <$M> + <$F> **A** = Small nod, short duration **B** = Small, multiple nods, longer duration than a **C** = Intense nod, short duration **D** = Intense, multiple nods, longer duration than c **E** = Multiple nods, combination of small & intense, long duration- usually intense nods fading to small nods **Speaker <$M> denoted in** grey**, <$F> in** green

Speaker	Details	'Nodder'	Backchannel (BC) Text	BC Type	Nod Type	Co-occurring Speech (if not as BC, text)	Nod Number	BC Number
<$M>	Nod	<$F>			A	restrict your data	2:1	
<$M>	Nod	<$F>			C	of your analysis	2:2	
<$M>	Nod	<$F>			A	to draw from	2:3	
<$F>	B-C and Nod	<$F>	Yeah	CNV	D	<$F> = Yeah <$M> = If they don't fit	2:4	2:1
<$M>	Nod	<$F>			C	looking at	2:5	
<$M>	Nod	<$F>			A	Quite a detailed	2:6	
<$F>	B-C and Nod	<$F>	Yeah	CON	C		2:7	2:2
<$F>	B-C and Nod	<$F>	No okay	ER	A		2:8	2:3
<$F>	B-C and Nod	<$F>	Yeah	CON	B	<$M> = Now at the beginning of July <$F> = Yeah	2:9	2:4
<$F>	B-C and Nod	<$F>	Yeah	CNV	A		2:10	2:5
<$M>	B-C		Erm	CON				1:1
<$M>	B-C and Nod	<$M>	Well yeah yeah	CNV	B	2= I think I would like to know <$E> laughs <\$E> what my <$X> data's | data is <\$X> gonna be as well so+ 1= Well yeah yeah	1:1	1:2
<$F>	Nod	<$M>			D	conditions listed some of them have masses of er	1:2	
<$M>	B-C		Right	IR				1:3
<$M>	B-C and Nod	<$M>	Right	IR	A		1:3	1:4
<$F>	B-C		Erm	CON				2:6
<$M>	B-C		Right	IR				1:5
<$F>	B-C		Is it	ER				2:7
<$F>	B-C and Nod	<$F>	Yeah	CON	A		2:11	2:8
<$M>	Nod	<$F>			C	A4 side	2:12	
<$M>	Nod	<$F>			A	six to seven thousand	2:13	
<$M>	B-C and Nod	<$M>	Okay	IR	C		1:4	1:6
<$M>	B-C		Oh wow right	ER				1:7
<$M>	B-C		Oh God	ER				1:8
<$M>	B-C		Yeah	CNV				1:9
<$M>	Nod	<$M>			B	So you're looking at	1:5	
<$F>	B-C and Nod	<$F>	Yeah	CNV	A		2:14	2:9

<$F>	Nod	<$F>			B	I think I am going to have to narrow it down	2:15	
<$F>	Nod	<$M>			C	narrow it down to a qualitative study	1:6	
<$M>	Nod	<$M>			A	Right	1:7	
<$F>	B-C	<$F>	Yeah erm	**CON**				2:10
<$M>	Nod	<$M>			A	or some sort of program	1:8	
<$M>	B-C		Erm yeah	**CON**				1:10
<$M>	B-C		Yeah something like that	**ER**				1:11
<$F>	B-C and Nod	<$F>	Yeah	**CNV**	C		2:16	2:11
<$M>	Nod	<$M>	Yeah		B		1:9	
<$F>	B-C and Nod	<$F>	Yeah	**CON**	C		2:17	2:12
<$M>	Nod	<$F>			A	sentence metaphors	2:18	
<$F>	Nod	<$F>			B	Yeah	2:19	
<$M>	B-C and Nod	<$M>	Right	**I R**	C		1:10	1:12
<$M>	B-C		Right yeah yeah	**I R**				1:13
<$M>	B-C and Nod	<$M>	Yeah yeah	**CNV**	A		1:11	1:14
<$F>	Nod	<$F>			A	verbs	2:20	
<$M>	B-C* 2 and Nod	<$M>	Right Yeah yeah	**I R** **CNV**	E	<$F> = how they're tagging it for them or whether they're just doing+ <$M> = Right. <$F> = +you know key word searches+ <$M> = Yeah yeah.	1:12	1:15 1:16
<$F>	Nod	<$F>			A	I think	2:21	
<$M>	B-C and Nod	<$M>	Right = Nod	**I R**	A	Right. Oh right yes. = B-C	1:13	1:17
<$F>	B-C and Nod	<$F>	Yeah I think so	**CNV**	C		2:22	2:13
<$M>	Nod	<$F>			C	even is	2:23	
<$M>	Nod	<$M>			C	But yeah you're right	1:14	
<$F>	B-C and Nod	<$F>	Yeah it would	**CNV**	C		2:24	2:14
<$F>	B-C and Nod	<$F>	Yeah *(first)*	**CON**	E	<$F> Yeah. <$M> +less data. <$F> Yeah I think I've already+ <$M> Okay. <$F> +decided that <$X> that's | that is <\$X> <$X> what's | what is	2:25	2:15
<$M>	B-C and Nod	<$M>	Okay	**I R**	C		1:15	1:18

< $M>	Nod	< $F>			A	criteria	2:26	
< $M>	Nod	< $F>			A	good examples	2:27	
< $F>	B-C		Yeah	CNV				2:16
< $F>	Nod	< $F>			B	some kind of principle	2:28	
< $M>	Nod	< $M>			C		1:16	
< $F>	B-C and Nod	< $F>	Yeah	CNV	B		2:29	2:17
< $M>	B-C and Nod	< $M>	Yeah yeah of course it is yeah	ER	A		1:17	1:19
< $M>	Nod	< $M>			A		1:18	
< $F>	B-C and Nod	< $F>	Yeah	CON	C		2:30	2:18
< $F>	B-C and Nod	< $F>	Yeah	CON	C		2:31	2:19
< $M>	B-C and Nod	< $M>	Okay = nod	ER	B	Erm yeah yeah okay	1:19	1:20
< $M>	B-C		Yeah right yeah	I R				1:21
< $M>	Nod	< $M>			A	Yeah yeah	1:20	
< $F>	Nod	< $M>			B	things at the moment	1:21	
< $M>	B-C		Right	I R				1:22
< $M>	B-C and Nod	< $M>	Uh-huh	CON	D		1:22	1:23
< $M>	B-C and Nod	< $M>	Right	I R	C	< $F>= educational purposes in+ < $M>= right	1:23	1:24
< $M>	B-C and Nod	< $M>	Yeah	CON	C	< $F>= concepts erm+ < $M>= yeah	1:24	1:25
< $F>	Nod	< $M>			C	Andrew Ortony book	1:25	
< $M>	B-C		Oh yeah	CNV				1:26
< $F>	Nod	< $M>			A	case studies and er medical text	1:26	
< $M>	B-C		Right	I R				1:27
< $F>	Nod	< $M>			C	war mapping even in	1:27	
< $F>	Nod	< $M>			B	which are	1:28	
< $M>	B-C		Yeah	CON				1:28
< $M>	B-C		Yeah yeah	CNV				1:29
< $M>	B-C and Nod	< $M>	Yeah	CNV	B		1:29	1:30
< $F>	Nod	< $M>			A	Pragmatics perspective of it	1:30	
< $F>	Nod	< $M>			C	You can get across a particular concept	1:31	
< $F>	Nod	< $M>			E	situation where it's perhaps quite difficult to explain	1:32	
< $M>	B-C		Right	I R				1:31
< $M>	B-C		Yeah	CNV				1:32

Appendix 4.1: A breakdown of the forms and functions of spoken backchannels in the case study data.

< $M>	B-C and Nod	< $M>	Okay	**I R**	A		1:33	1:33
< $F>	Nod	< $F>			C	Yeah so where it	2:32	
< $F>	Nod	< $M>			A	to approach er these patient narratives	1:34	
< $F>	Nod	< $F>			E	patient narratives	2:33	
< $M>	B-C		Yeah	**CON**				1:34
< $F>	Nod	< $M>			B	of saying how it is	1:35	
< $M>	B-C		Yeah	**CON**				1:35
< $F>	B-C and Nod	< $F>	Yes	**CNV**	C		2:34	2:20
< $M>	Nod	< $F>			A	contrastive study	2:35	
< $M>	Nod	< $F>			A	talk about stuff	2:36	
< $M>	Nod	< $F>			A	talk about stuff	2:37	
< $M>	Nod	< $F>			C	two sets	2:38	
< $M>	Nod	< $F>			A	one hand	2:39	
< $M>	Nod	< $F>			A	Do you see what I mean?	2:40	
< $M>	B-C and Nod	< $M>	Right yeah	**I R**	C		1:36	1:36
< $F>	Nod	< $F>			C	Yes	2:41	
< $M>	B-C * 3 and Nod	< $M>	Oh right yeah Yeah Yeah	**ER** **CON** **CON**	E	< $M> = Oh right yeah < $F> = +medical practitioners anyway to explain+ < $M> = Yeah < $F> = +what they have been going through. So that's+ < $M> = Yeah < $F> = +sort of the	1:37	1:37 1:38 1:39
< $F>	B-C and Nod	< $F>	Yeah	**CNV**	C		2:42	2:21
< $F>	B-C and Nod	< $F>	Yeah	**CNV**	C		2:43	2:22
< $F>	Nod	< $F>			B	yeah there's quite a lot of papers on that	2:44	
< $M>	Nod	< $M>			B	yeah I'm conscious	1:38	
< $F>	B-C and Nod	< $F>	Yeah	**CON**	C		2:45	2:23
< $M>	Nod	< $F>			A	you're the expert yeah.	2:46	
< $M>	Nod	< $F>			A	either exclude either one	2:47	
< $M>	Nod	< $F>			A	study	2:48	
< $M>	B-C and Nod	< $M>	Oh yeah.	**ER**	E		1:39	1:40
< $F>	B-C and Nod	< $F>	Yeah	**CON**	A		2:49	2:24
< $M>	Nod	< $F>			C	doctor to somebody else	2:50	
< $M>	Nod	< $F>			C	to somebody else	2:51	

Appendix 4.1: A breakdown of the forms and functions of spoken backchannels in the case study data.

<$M>	Nod	<$F>			C	accommodating towards each other	2:52	
<$M>	Nod	<$F>			A	before the <$H> setting	2:53	
<$M>	Nod	<$M>			A	+you see what I mean+	1:40	
<$F>	B-C and Nod	<$F>	Yeah	**CNV**	A		2:54	2:25
<$M>	Nod	<$F>			C	data that you've got	2:55	
<$F>	B-C (2) and Nod (1)	<$M>	Okay	**I R**	A		1:41	2:26
<$M>	Nod	<$F>			C	not to each other	2:56	
<$F>	Nod	<$F>			B	Just <\$=> Yeah in contrast+	2:57	
<$F>	Nod	<$F>			C	Yeah that would be quite interesting	2:58	
<$M>	B-C		Mm	**CON**				1:41
<$F>	Nod	<$M>			B	choosing those particular domains	1:42	
<$M>	B-C		Yeah	**CNV**				1:42
<$F>	B-C		Erm.	**CON**				2:27
<$M>	Nod	<$M>			E	Yeah Yeah.	1:43	
<$M>	Nod	<$F>			A	discoursey pragmatic thing.	2:59	
<$M>	Nod	<$F>			A	interferes with it all	2:60	
<$M>	Nod	<$F>			A	other factors	2:61	
<$F>	Nod	<$F>			E	<$F>= Yeah well+ <$M>= +want to talk about.	2:62	
<$M>	B-C and Nod	<$M>	Yeah yeah.	**CNV**	C		1:44	1:43
<$M>	Nod	<$F>			A	Yeah yeah.	2:63	
<$M>	B-C and Nod	<$M>	Yeah	**CON**	B		1:45	1:44
<$M>	B-C and Nod	<$M>	Yeah	**CON**	A	<$M>= Yeah+ <$F>= +it would be far too big to+	1:46	1:45
<$F>	B-C and Nod	<$F>	Yeah	**CNV**	B		2:64	2:28
<$M>	Nod	<$F>			C	don't forget these ideas	2:65	
<$M>	Nod	<$F>			B	bottom drawer ideas	2:66	
<$M>	Nod	<$F>			A	have you got your article for this	2:67	
<$M>	Nod	<$F>			A	do you want to come to this	2:68	
<$M>	Nod	<$F>			C	never had time to word	2:69	
<$M>	Nod	<$F>			A	then you do those	2:70	

Appendix 4.1: A breakdown of the forms and functions of spoken backchannels in the case study data.

Appendix 4.2: Frequency counts of individual spoken backchannel forms in the case study data (co-occurring with nods and without nods).

Total number of spoken backchannels used: 73

- **erm (3)**
  - Without Nod → <$M> → CON 1
  - → <$F> → CON 2
- **erm yeah** → Without Nod → <$M> → CON 1
- **is it** → Without Nod → <$F> → ER 1
- **mm** → Without Nod → <$M> → CON 1
- **no okay** → Nod → <$F> → TYPE A → IR 1
- **oh god** → Without Nod → <$M> → ER 1
- **oh right yeah** → Nod → <$M> → ER 1
- **oh wow right** → Without Nod → <$M> → ER 1
- **oh yeah (2)**
  - Without Nod → <$M> → CNV 1
  - Nod → <$M> → TYPE E → ER 1
- **okay (5)**
  - Nod (3) → <$M> → TYPE A → IR 1
    - TYPE B → ER 1
    - TYPE C → IR 1
  - → <$F> → TYPE A → IR 1
  - Without Nod → <$F> → IR 1
- **right (10)**
  - Without Nod → <$M> → IR 5
  - Nod (5) → <$M> → TYPE A → IR 2
    - TYPE C → IR 2
    - TYPE E → ER 1
- **right yeah** → Nod → <$M> → TYPE C → IR 1
- **right yeah yeah** → Without Nod → <$M> → IR 1
- **Uh-huh** → Nod → <$M> → TYPE D → CON 1
- **Well yeah yeah** → Nod → <$M> → TYPE B → CNV 1
- **yeah (31)**
  - Without Nod (7)
    - <$M> (6) → CON 3
      - CNV 3
    - <$F> (1) → CNV 1
  - Nod (24)
    - <$M> (6) → TYPE A → CON 1
      - TYPE B (2) → CON 1 / CNV 1
      - TYPE C → CNV 1
      - TYPE E → CON 2
    - <$F> (18)
      - TYPE A (5) → CON 2 / CNV 3
      - TYPE B (3) → CON 1 / CNV 2
      - TYPE C (8) → CON 5 / CNV 3
      - TYPE D → CNV 1
      - TYPE E → CON 1
- **yeah erm** → Without Nod → <$F> → CON 1
- **Yeah I think so** → Nod → <$F> → TYPE C → CNV 1
- **Yeah it would** → Nod → <$F> → TYPE C → CNV 1
- **yeah right yeah** → Without Nod → <$M> → IR 1
- **yeah something like that** → Without Nod → <$M> → ER 1
- **yeah yeah (4)**
  - Without Nod → <$M> → CNV 1
  - Nod (3) → <$M> → TYPE A → CNV 1
    - TYPE C → CNV 1
    - TYPE E → CNV 1
- **yeah yeah of course it is yeah** → Nod → <$M> → TYPE A → ER 1
- **yes** → Nod → <$F> → TYPE C → CNV 1

KEY:

Symbol	Meaning
<$M>	Supervisor
<$F>	Supervisee
A	Type A nod
B	Type B nod
C	Type C nod
D	Type D nod
E	Type E nod
CON	Continuer
CNV	Convergence Token
ER	Engaged Response Token
IR	Information Receipt Token
2	Frequency of occurrence

Appendix 4.3: Frequency counts of individual spoken backchannel functions in the case study data (co-occurring with nods and without nods).
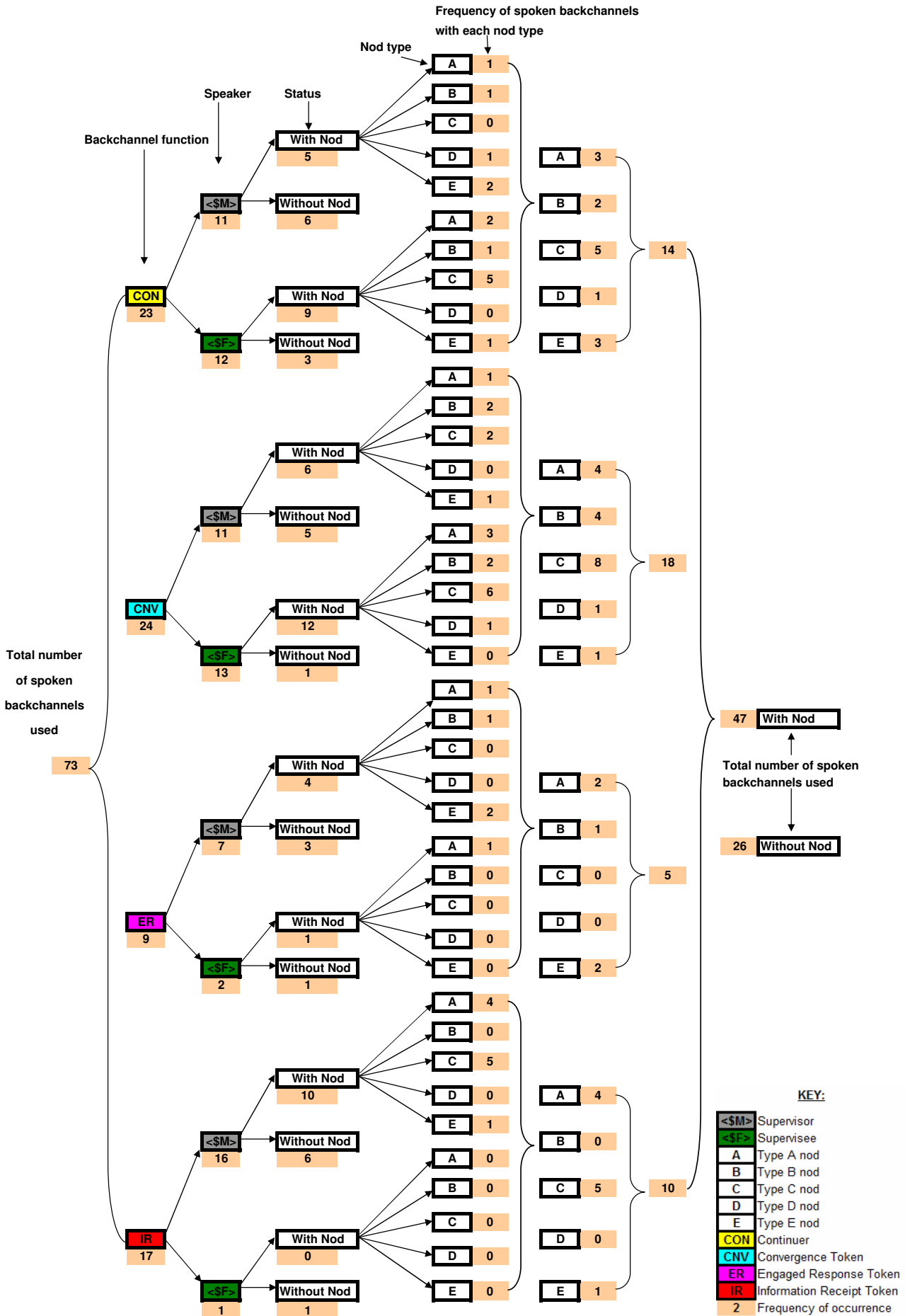
Appendix 4.4: Frequencies of backchanneling head nod 'types' in the case study data.
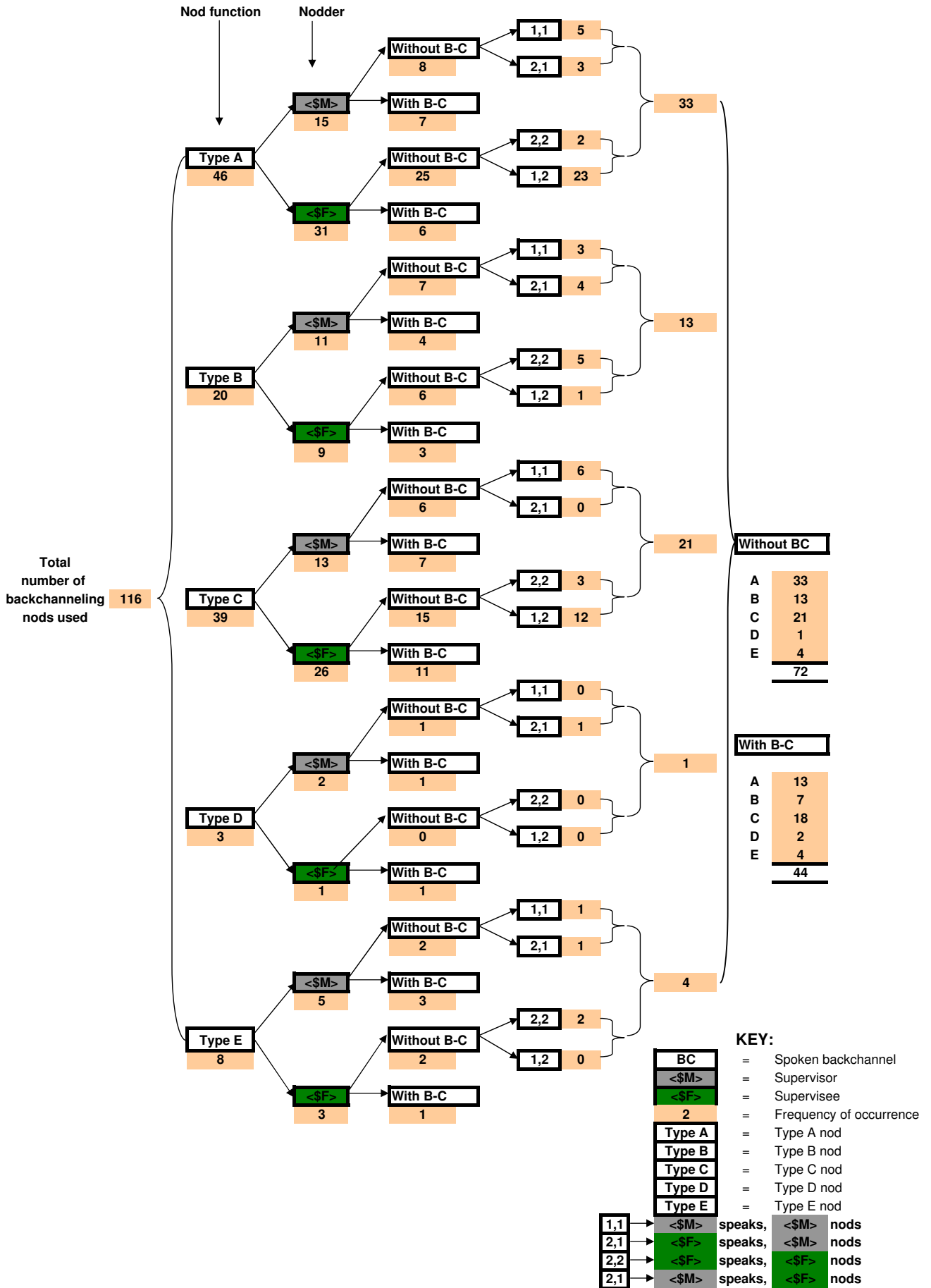
Appendix 5.1: The most intense peaks and troughs tracked for S03MF.M in the case study excerpt.

Table 1: Intense head peaks

PEAKS: Code	Frame(s)	Y Axis Value Min.	Max.	Nod?	No.
	127 to 184	35	36	y	N/A
	278 to 307	35	51	n	
p1	354 to 381	35		n	-
p2	428	35		n	-
p3	720	35		n	-
p4	793	35		n	-
p5	804 to 817	38	47	n	-
p6	1075	35		n	-
p7	2895 to 2919	35		n	-
p8	3039 to 3045	35		y	6C
p9	3310 to 3313	35		n	-
p10	3682 to 3692	35	56	y	10C
p11	3829	35		n	-
p12	4620 to 4621	35		y	11A
p13	4722 to 4732	36	70	y	12E
p14	5094	35		n	-
p15	5226 to 5227	35		y	15C
p16	5303 to 5322	35	41	n	-
p17	6072 to 6075	35	36	y	17A
p18	6197 to 6209	36	41	y	18A
p19	6634 to 6650	35	59	n	-
p20	6739 to 6751	35	40	n	-
p21	6843 to 6857	35	39	n	-
p22	6918 to 6924	35	37	y	22D
p23	7186 to 7190	35		y	25C
p24	7624	35		n	-
p25	7750 to 7754	35	37	y	28B
p26	8160 to 8161	37	40	n	-
p27	8348 to 8363	36	41	n	-
p28	8397 to 8398	35		n	-
p29	9019 to 9023	37	40	n	-
p30	9109 to 9111	35	36	n	-
p31	9302 to 9322	35		y	34E
p32	10045 to 10054	37	46	n	-
p33	10926 to 10966	35	42	n	-
p34	11481 to 11490	36	50	n	-
p35	11571	35		y-nbc	-
p36	11598 to 11629	35	41	n	-
p37	11723 to 11726	35	37	n	-
p38	12449 to 12473	35	44	n	-
p39	12937 to 12947	35	40	n	-
p40	13350 to 13973	35	39	n	-

Average = 30.136
S.D. = 2.5223
2*S.D. = 5.0446
Exploring the range: 25 ≤ x ≥ 35
Combined Clusters: 25 frames ≤≥ of each other (p/t)

Table 2: Intense head troughs

TROUGHS: Code	Frame(s)	Y Axis Value Min.	Max.	Nod?	No.
t1	978 to 984	24	25	n	-
t2	1388 to 1390	25		n	-
t3	2065 to 2091	20	24	y	3A
t4	3668 to 3681	17	25	y	10C
	3713 to 3716	25		n	-
t5	3838 to 3839	25		n	-
t6	3956 to 3966	25		n	-
t7	4711 to 4721	3	25	y	12E
t8	5928 to 5946	20	25	n	-
t9	6033	25		n	-
t10	6619 to 6629	18	24	n	-
t11	6759 to 6794	23	25	n	-
t12	6818 to 6837	25		n	-
t13	7596 to 7617	22	25	n	-
t14	10381 to 10384	25		n	-
t15	10917 to 10921	22	23	n	-
t16	11475 to 11477	23		n	-
t17	12508 to 12515	24	25	n	-
t18	13301 to 13306	23	25	n	-
t19	14492	25		n	-
t20	15146 to 15168	25		n	-

Table 3: Combined intense peaks and troughs

Code	Frame(s)	Y Axis Value Min.	Max.	Up/Down?	Nod No.
1	3682 to 3716	25	35	y|y	10C
2	3829 to 3839	25	35	n|n	-
3	4711 to 4732	3	70	y|y	12E
4	6619 to 6650	18	59	n|n	-
5	6739 to 6794	23	40	n|n	-
6	6818 to 6857	25	39	n|n	-
7	7596 to 7624	22	35	n|n	-
8	10917 to 10966	22	42	n|n	-
9	11475 to 11490	23	50	n|n	-
10	12449 to 12515	24	44	n|n	-

KEY:

p	=	Peaks (data under focus)
t	=	Troughs (data under focus)
	=	Frames part of a peak-trough cluster
	=	Frames part of a peak-trough cluster
y	=	Yes (movement is a head nod)
n	=	No (movement is not a head nod)
	=	frame clusters aligned as a result of peak-trough clusters
	=	Data preceding the Case Study
nbc	=	Nod, but not a backchanneling nod
No.	=	Nod number and coded nod type (as ascribed in chapter 4)

Appendix 5.2: 'Medium-sized' peaks and troughs tracked for S03MF.M in the case study excerpt.

Table 1: Medium-sized head peaks

PEAKS:			
Code	**Frame(s)**	**Nod?**	**No.**
	38	n	
	124 to 148	n	
	176 to 182	n	N/A
	266	n	
	335 to 389	n	
sp1	427 to 430	n	-
sp2	581 to 605	n	-
sp3	706 to 733	n	-
sp4	792 to 803	y-nbc	-
sp5	1071 to 1076	n	-
sp6	1180	n	-
sp7	1424	n	-
sp8	2820 to 2821	n	-
sp9	2896 to 2931	y	5B
sp10	3041 to 3043	y	6C
sp11	3077 to 3080	y	7A
sp12	3127	y	8A
sp13	3264 to 3265	n	-
sp14	3280 to 3284	n	-
sp15	3314	n	-
sp16	3412 to 3413	n	-
sp17	3490 to 3496	n	-
sp18	3828 to 3830	n	-
sp19	4622	y	13A
sp20	5056 to 5103	y	14A
sp21	5217 to 5228	n	-
sp22	6068 to 6076	n	-
sp23	6124	y	17A
sp24	6632 to 6633	n	-
sp25	6842	n	-
sp26	6901 to 6928	y	22D
sp27	7188 to 7193	y	25C
sp28	7623	n	-
sp29	7865 to 7867	y	29B
sp30	8070 to 8088	n	-
sp31	8159	n	-
sp32	8347	n	-
sp33	8395 to 8399	n	-
sp34	9240 to 9255	y	34E
sp35	9298 to 9436	y	35B
sp36	10044	n	-
sp37	10924 to 10944	n	-
sp38	11210 to 11212	y	38B
sp39	11480	n	-
sp40	11560 to 11611	n	-
sp41	11660 to 11661	n	-
sp42	11722	n	-
sp43	12066	y	39B
sp44	12220 to 12222	n	-
sp45	12312 to 12319	n	-
sp46	12359 to 12360	n	-
sp47	12424 to 12457	n	-
sp48	12935 to 12936	n	-
sp49	13457 to 13458	y	43E
sp50	13784	n	-
sp51	13974	n	-
sp52	14587	n	-
sp58	14626	n	-

Average = 30.136
Exploring the axis values: 26, 34
Combined Clusters: 25 frames ≤≥ of each other

Table 2: Medium-sized head troughs

TROUGHS:			
Code	**Frame(s)**	**Nod?**	**No.**
	118 to 119	n	N/A
st1	587 to 593	n	-
st2	985	n	-
st3	1387	n	-
st4	1742 to 1747	y	1B
st5	3529 to 3544	y	9B
st6	3693 to 3696	n	-
st7	3711 to 3736	n	-
st8	3840 to 3843	n	-
st9	3937 to 4012	n	-
st10	4061 to 4078	n	-
st11	4197 to 4229	y	11A
st12	5483 to 5485	n	-
st13	5927 to 5948	n	-
st14	6031 to 6034	n	-
st15	6466 to 6468	y	20A
st16	6499	n	-
st17	6719	n	-
st18	6758	n	-
st19	6785 to 6836	n	-
st20	7591 to 7598	n	-
st21	7618	n	-
st22	7704 to 7705	y	27C
st23	7987	n	-
st24	8991 to 9006	n	-
st25	10380 to 10390	n	-
st26	10760 to 10765	n	-
st27	10916	n	-
st28	11474	n	-
st29	12105 to 12156	n	-
st30	12506 to 12522	n	-
st31	12763 to 12764	y	40A
st32	12816 to 12817	n	-
st33	12865 to 12891	y	41A
st34	13300 to 13307	n	-
st35	13328 to 13329	n	-
st36	13856	n	-
st37	14483 to 14523	n	-
st38	15138 to 15220	n	-
st39	15281	n	-
st40	15296	n	-

Table 3: Combined intense peaks and troughs

Code	Frame(s)	Up/ Down?	Nod No.
1	581 to 605	n|n	-
2	3828 to 3843	n|n	-
3	6785 to 6842	n|n	-
4	7618 to 7623	n|n	-
5	10916 to 10944	n|n	-
6	11474 to 11480	n|n	-

KEY:

sp	=	Medium-sized peaks
st	=	Medium-sized troughs
	=	Frames part of a peak-trough cluster
	=	Frames part of a peak-trough cluster
y	=	Yes (movement is a head nod)
n	=	No (movement is not a head nod)
	=	Data preceding the Case Study
nbc	=	Nod, but not a backchanneling nod
No.	=	Nod number and coded nod type (as ascribed in chapter 4)

Table 1: Intense and medium-sized head peaks

Code		Frame(s)	Y Axis Value Min.	Max.	Nod?	No.
	sp	38	34		n	
	sp	124 to 148	34		n	
	p	127 to 184	35	36	y	
	sp	176 to 182	34		n	N/A
	sp	266	34		n	
	p	278 to 307	35	51	n	
	sp	335 to 389	34		n	
cp1	p1	354 to 381	35		n	-
cp2	sp1	427 to 430	34		n	-
	p2	428	35		n	-
cp3	sp2	581 to 605	34		n	-
cp4	sp3	706 to 733	34		n	-
	p3	720	35		n	-
cp5	sp4	792 to 803	34		y	nbc
	p4	793	35		n	-
	p5	804 to 817	38	47	n	-
cp6	sp5	1071 to 1076	34		n	-
	p6	1075	35		n	-
	sp6	1180	34		n	-
cp7	p7	1424	34		n	-
cp8	sp8	2820 to 2821	34		n	-
cp9	p7	2895 to 2919	35		n	-
	sp9	2896 to 2931	34		y	5B
	p8	3039 to 3045	35		y	6C
	sp10	3041 to 3043	34		y	6C
cp10	sp11	3077 to 3080	34		y	7A
cp11	sp12	3127	34		y	8A
cp12	sp13	3264 to 3265	34		n	-
	sp14	3280 to 3284	34		n	-
cp13	p9	3310 to 3313	35		n	-
	sp15	3314	34		n	-
cp14	sp16	3412 to 3413	34		n	-
cp15	sp17	3490 to 3496	34		n	-
cp16	p10	3682 to 3692	35	56	y	10C
cp17	sp18	3828 to 3830	34		n	-
	p11	3829	35		n	-
cp18	p12	4620 to 4621	35		y	11A
	sp19	4622	34		y	11A
cp19	p13	4722 to 4732	36	70	y	12E
cp20	sp20	5056 to 5103	34		y	14A
	p14	5094	35		n	-
cp21	sp21	5217 to 5228	34		n	-
	p15	5226 to 5227	35		y	15C
cp22	p16	5303 to 5322	35	41	n	-
cp23	sp22	6068 to 6076	34		n	-
	p17	6072 to 6075	35	36	y	17A
cp24	sp23	6124	34		y	17A
cp25	p18	6197 to 6209	36	41	y	18A
cp26	sp24	6632 to 6633	34		n	-
	p19	6634 to 6650	35	59	n	-
cp27	p20	6739 to 6751	35	40	n	-
cp28	sp25	6842	34		n	-
	p21	6843 to 6857	35	39	n	-
cp29	sp26	6901 to 6928	34		y	22D
	p22	6918 to 6924	35	37	y	22D
cp30	p23	7186 to 7190	35		y	25C
	sp27	7188 to 7193	34		y	25C
cp31	sp28	7623	34		n	-
	p24	7624	35		n	-
cp32	p25	7750 to 7754	35	37	y	28B
cp33	p29	7865 to 7867	34		y	29B
cp34	sp30	8070 to 8088	34		n	-
cp35	sp31	8159	34		n	-
	p26	8160 to 8161	37	40	n	-
cp36	sp32	8347	34		n	-
	p27	8348 to 8363	36	41	n	-
cp37	sp33	8395 to 8399	34		n	-
	p28	8397 to 8398	35		n	-
cp38	p29	9019 to 9023	37	40	n	-
cp39	p30	9109 to 9111	35	36	n	-
cp40	sp34	9240 to 9255	34		y	34E
cp41	sp35	9298 to 9436	34		y	35B
	p31	9302 to 9322	35		y	34E
cp42	sp36	10044	34		n	-
	p32	10045 to 10054	37	46	n	-
cp43	sp37	10924 to 10944	34		n	-
	p33	10926 to 10966	35	42	n	-
cp44	sp38	11210 to 11212	34		y	38B
cp45	sp39	11480	34		n	-
	p34	11481 to 11490	36	50	n	-
cp46	sp40	11560 to 11611	34		n	-
	p35	11571	35		y	nbc
	p36	11598 to 11629	35	41	n	-
cp46	sp41	11660 to 11661	34		n	-
cp47	sp42	11722	34		n	-
	p37	11723 to 11726	35	37	n	-
cp48	sp43	12066	34		y	39E
cp49	sp44	12220 to 12222	34		n	-
cp50	sp45	12312 to 12319	34		n	-
cp51	sp46	12359 to 12360	34		n	-
cp52	sp47	12424 to 12457	34		n	-
	p38	12449 to 12473	35	44	n	-
cp53	sp48	12935 to 12936	34		n	-
	p39	12937 to 12947	35	40	n	-
cp54	p40	13350 to 13973	35	39	n	-
	sp49	13457 to 13458	34		y	43E
cp55	sp50	13784	34		n	-
cp56	sp51	13974	34		n	-
cp57	sp52	14587	34		n	-
cp58	sp58	14626	34		n	-

Table 2: Intense and medium-sized head troughs

Code		Frame(s)	Y Axis Value Min.	Max.	Nod?	No.
	st	118 to 119	26		n	-
ct1	st1	587 to 593	26		n	-
ct2	t1	978 to 984	24	25	n	-
	st2	985	26		n	-
ct3	st3	1387	26		n	-
	t2	1388 to 1390	25		n	-
ct4	st4	1742 to 1747	26		y	1B
ct5	t3	2065 to 2091	20	24	n	3A
ct6	st5	3529 to 3544	26		y	9B
ct7	t4	3668 to 3681	17	25	y	10C
	st6	3693 to 3696	26		n	-
	st7	3711 to 3736	26		n	-
	t5	3713 to 3716	25		n	-
ct8	t6	3838 to 3839	25		n	-
	st8	3840 to 3843	26		n	-
ct9	st9	3937 to 4012	26		n	-
	t7	3956 to 3966	25		n	-
ct10	st10	4061 to 4078	26		n	-
ct11	st11	4197 to 4229	26		y	11A
ct12	t8	4711 to 4721	3	25	y	12E
ct13	st12	5483 to 5485	26		n	-
ct14	st13	5927 to 5948	26		n	-
	t9	5928 to 5948	25		n	-
ct15	st14	6031 to 6034	26		n	-
	t10	6033	25		n	-
ct16	st15	6466 to 6468	26		y	20A
ct17	st16	6499	26		n	-
ct18	t11	6619 to 6629	18	24	n	-
ct19	st17	6719	26		n	-
	st18	6758	26		n	-
ct20	t12	6759 to 6794	23	25	n	-
	st19	6785 to 6836	26		n	-
	t13	6818 to 6837	25		n	-
ct21	st20	7591 to 7598	26		n	-
	t14	7596 to 7617	22	25	n	-
	st21	7618	26		n	-
ct22	st22	7704 to 7705	26		y	27C
ct23	st23	7987	26		n	-
ct24	st24	8991 to 9006	26		n	-
ct25	st25	10380 to 10390	26		n	-
	t15	10381 to 10384	25		n	-
ct26	st26	10760 to 10765	26		n	-
ct27	st27	10916	26		n	-
	t16	10917 to 10921	22	23	n	-
ct28	st28	11474	26		n	-
	t17	11475 to 11477	23		n	-
ct29	st29	12105 to 12156	26		n	-
ct30	st30	12506 to 12522	26		n	-
	t18	12508 to 12515	24	25	n	-
ct31	st31	12763 to 12764	26		y	40A
ct32	st32	12816 to 12817	26		n	-
ct33	st33	12865 to 12891	26		y	41A
ct34	st34	13300 to 13307	26		n	-
	t19	13301 to 13306	23	25	n	-
ct35	st35	13328 to 13329	26		n	-
ct36	st36	13856	26		n	-
ct37	st37	14483 to 14523	26		n	-
	t20	14492	25		n	-
ct38	st38	15138 to 15220	26		n	-
	t21	15146 to 15168	25		n	-
ct39	st39	15281	26		n	-
	st40	15296	26		n	-

Table 3: Combined peaks and troughs for intense and medium-sized nods

Code	Frame(s)	Y Axis Value Min.	Max.	B/M Peaks?	B/M Troughs?	Nod No.
c1	581 to 605	26	34	n|sp2	n|st1	n|n|n|n
c2	3668 to 3696	17	56	p10|n	t4|st6	10C|n|10C|n
c4	3828 to 3843	25	34	p11|sp18	t6|st8	n|n|n|n
c5	4711 to 4732	3	70	p13|n	t8|n	12E|n|12E|n
c6	6619 to 6650	18	59	p19|sp24	t11|n	n|n|n|n
c7	6719 to 6836	23	40	p20|n	t12|st17/8/9	n|n|n|n
c8	6818 to 6857	25	39	p21|sp25	t13|n	n|n|n|n
c9	7591 to 7624	22	35	p24|sp28	t14|st20/1	n|n|n|n
c10	8991 to 9023	26	40	p29|n	n|st24	n|n|n|n
c11	10916 to 10966	22	42	p33|sp37	t16|st27	n|n|n|n
c12	11474 to 11490	23	50	p34|sp39	t17|st28	n|n|n|n
c13	13300 to 13973	23	39	p40|n	t19|st34/5	n|n|n|n

KEY:

p	=	Intense peaks
t	=	Intense troughs
sp	=	Medium-sized peaks
st	=	Medium-sized troughs
(yellow)	=	Frames part of a peak-trough cluster
(grey)	=	Frames part of a peak-trough cluster
y	=	Yes (movement is a head nod)
n	=	No (movement is not a head nod)
(hatched)	=	Frame clusters aligned as a result of peak-trough clusters
(blue)	=	Points where intense and medium-sized peaks and troughs cluster (within 25 frames)
(cyan)	=	Data preceding the Case Study
cp	=	Peaks across intense and/or medium-size movements (c = combined)
ct	=	Troughs across intense and/or medium-size movements (c = combined)
c	=	Combined peak and trough across the intense and / or medium-sized movements
nbc	=	Nod, but not a backchanneling nod
No.	=	Nod number and coded nod type (as ascribed in chapter 4)

Average = 30.136
S.D. = 2.5223
2*S.D. = 5.0446
Exploring the range: $26 \leq x \geq 34$
Combined Clusters: 25 frames $\leq\geq$ of each other

Appendix 5.3: Combining the most intense and 'medium-sized' head peaks and troughs for S03MF.M, as seen across the case study excerpt (for a closer analysis of clusters of head movement).

Appendix 5.4: The most intense peaks and troughs tracked for S03MF.F in the case study excerpt.

Table 1: Intense head peaks

PEAKS:		Y Axis Value			
Code	Frame(s)	Min.	Max.	Nod?	No.
	18 to 21	30		n	N/A
	86	30		n	
	138 to 149	30	38	y	
	179 to 188	30	41	n	
p1	1721	30		n	-
p2	1966 to 1969	30	32	n	-
p3	2584 to 2599	30	31	y- nbc	-
p4	2982	30		y	15B
p5	3165 to 3166	30		n	-
p6	3372 to 3396	30	32	n	-
p7	3973 to 3980	30		y	17C
p8	3998 to 4012	30	32	y	18A
p9	4033 to 4060	30	31	n	-
p10	4127 to 4132	30		n	-
p11	4154 to 4167	30	32	n	-
p12	4192 to 4193	30		n	-
p13	4207 to 4208	30		n	-
p14	4284 to 4287	30	31	n	-
p15	4326 to 4333	30	31	n	-
p16	5737 to 5753	30	32	y	29B
p17	6354 to 6363	30	32	n	-
p18	6813 to 6837	30	40	n	-
p19	7000 to 7002	30		n	-
p20	7023 to 7025	30		n	-
p21	7160 to 7162	30		n	-
p22	7244 to 7245	30		n	-
p23	7305 to 7310	30		n	-
p24	8435 to 8437	31		n	-
p25	8688 to 8693	30		n	-
p26	9029 to 9030	30		y	33A
p27	9219 to 9220	30		n	-
p28	9266 to 9269	30		n	-
p29	9271 to 9278	30		n	-
p30	9336 to 9338	30		n	-
p31	9351 to 9353	30		n	-
p32	10378	30	31	y	41C
p33	10435 to 10437	32		n	-
	10475 to 10493	30	32	y	42C
p34	10686 to 10699	30		y	43C
p35	12021 to 12023	30		n	-
p36	12351 to 12365	30		y	50C/51C
p37	12713 to 12725	30		y	54A
p38	12751 to 12775	30	31	y	55C
p39	13050 to 13062	30	32	y	57B
p40	13214 to 13216	30	31	y	58C
p41	13897	30		y	62E
p42	14028	30	32	n	-

Average = 25.33632
S.D. = 2.24729
2*S.D. = 4.49458
Exploring the range: 20/1 ≤ x ≥ 30
Combined Clusters: 25 frames ≤≥ of each

Table 2: Intense head troughs

TROUGHS:		Y Axis Value			
Code	Frame(s)	Min.	Max.	Nod?	No.
t1	470 to 479	17	19	y	2C
t2	1318 to 1338	18	20	n	-
t3	1455 to 1456	20		y	9B
t4	2705	20		n	-
t5	3347 to 3348	20		n	-
t6	3448 to 3451	20		n	-
t7	3720 to 3723	20		y	16C
t8	4377 to 4384	19	20	n	-
t9	4526	20		n	-
t10	5383 to 5384	20		n	-
t11	5947 to 6001	20		n	-
t12	6635 to 6645	18	20	n	-
t13	7451 to 7464	18	20	n	-
t14	8265	20		n	-
t15	8546 to 8557	19	20	n	-
t16	8959 to 8961	20		n	-
t17	9145 to 9156	17	20	y	32C
t18	10389 to 10400	18	20	y	41C
t19	10450 to 10467	18	20	y	42C
t20	11941 to 11944	20		y-nbc	-
t21	13023 to 13046	16	20	y	57B
t22	13918 to 13944	19	20	y	62E
t23	13988 to 13992	20		n	-
t24	14035 to 14037	19	20	n	-
	14050 to 14061	18	20	n	-
	14086 to 14103	18	20	n	-

Table 3: Combined medium-sized peaks and troughs

		Y Axis Value		Up/	Nod
Code	Frame(s)	Min.	Max.	Down?	No.
1	3372 to 3451	20	32	n|n	-
2	10378 to 10400	18	31	y|y	41C
3	10435 to 10493	18	32	ny|y	42C
4	13023 to 13062	16	32	y|y	57B
5	13897 to 13944	19	30	y|y	62E
6	14028 to 14103	22	32	n|n	-

KEY:

p	=	Peaks (data under focus)
t	=	Troughs (data under focus)
(yellow)	=	Frames part of a peak-trough cluster
(grey)	=	Frames part of a peak-trough cluster
y	=	Yes (movement is a head nod)
n	=	No (movement is not a head nod)
(dotted)	=	frame clusters aligned as a result of peak-trough clusters
(cyan)	=	Data preceding the Case Study
nbc	=	Nod, but not a backchanneling nod
No.	=	Nod number and coded nod type (as ascribed in chapter 4)

Appendix 5.5: 'Medium-sized' peaks and troughs tracked for S03MF.F in the case study excerpt.

Table 1: Medium-sized head peaks

PEAKS:				
Code	**Frame(s)**		**Nod?**	**No.**
	22 to 92		n	N/A
sp1	238 to	243	n	-
sp2	304 to	381	n	-
sp3	863 to	871	y	4D
sp4	898 to	918	y	5C
sp5	1153 to	1154	y	6A
sp6	1272 to	1273	y	8A
sp7	1700		n	-
sp8	1790 to	1796	y	-
sp9	1949 to	1952	n	-
sp10	2438 to	2444	y	11A
sp11	2769 to	2776	n	-
sp12	2826		n	-
sp13	2974 to	2983	y	14B
sp14	3163 to	3171	n	-
sp15	3270 to	3277	n	-
sp16	3362 to	3371	y	16C
sp17	3637 to	3655	n	-
sp18	3891 to	3892	n	-
sp19	3948 to	4234	y	19B
sp20	4283 to	4289	n	-
sp21	4297 to	4298	n	-
sp22	4674		y	22C
sp23	4837 to	4908	n	-
sp24	5735 to	5736	y	28B
sp25	6352 to	6365	n	-
sp26	6779 to	6812	n	-
sp27	6855 to	7026	n	-
sp28	7163 to	7164	n	-
sp29	7243 to	7249	n	-
sp30	7311		n	-
sp31	7558 to	7564	n	-
sp32	7705 to	7706	n	-
sp33	8377		n	-
sp34	8428 to	8434	n	-
sp35	8521		n	-
sp36	8685 to	8730	n	-
sp37	8778 to	8780	n	-
sp38	9023 to	9033	y	32C
sp39	9134		n	-
sp40	9218 to	9221	n	-
sp41	9249 to	9281	n	-
sp42	9318 to	9383	n	-
sp43	10286 to	10302	n	-
sp44	10355 to	10382	n	-
sp45	10438		n	-
sp46	10474 to	10492	n	-
sp47	10685 to	10746	n	-
sp48	11042 to	11048	y	44B
sp49	11370 to	11374	y	45C
sp50	11392 to	11401	n	-
sp51	12007 to	12024	n	-
sp52	12218 to	12222	y	49B
sp53	12344 to	12366	y	50C
sp54	12711 to	12750	n	-
sp55	13049 to	13084	y	57B
sp56	13110 to	13115	y	58C
sp57	13131 to	13141	n	-
sp58	13204 to	13218	n	-
sp59	13887 to	13903	y	62E
sp60	14024 to	14029	n	-
sp61	14135 to	14140	y	63A
sp62	14427 to	14451	n	-
sp63	15210 to	15230	y	70A

Average = 25.33632

Exploring the axis values: 21, 29

Table 2: Medium-sized head troughs

TROUGHS:				
Code	**Frame(s)**		**Nod?**	**No.**
st1	466 to	469	y	2C
st2	708 to	747	y	3A
st3	1310 to	1347	n	-
st4	1454 to	1457	y	9B
st5	1628 to	1632	y	10A
st6	1658 to	1661	n	-
st7	2698 to	2709	n	-
st8	3345 to	3351	y	16C
st9	3450 to	3455	n	-
st10	3693 to	3768	y	17C
st11	4379 to	4385	n	-
st12	4426 to	4427	n	-
st13	4520 to	4529	y	21A
st14	5371 to	5388	n	-
st15	5901 to	5904	n	-
st16	5934 to	5946	y	29B
st17	5993 to	5998	n	-
st18	6376 to	6395	n	-
st19	6639		n	-
st20	6724		n	-
st21	7369 to	7386	n	-
st22	7419 to	7458	n	-
st23	8248 to	8264	n	-
st24	8451 to	8544	n	-
st25	8864 to	8958	n	-
st26	9144 to	9150	n	-
st27	10031 to	10116	y	40A
st28	10388 to	10411	y	41C
st29	10449 to	10468	n	-
st30	10604 to	10620	n	-
st31	10927 to	10936	y	43C
st32	11580		y	47B
st33	11881		n	-
st34	11939 to	11940	n	-
st35	11978 to	11979	n	-
st36	12043 to	12051	n	-
st37	13022		y	57B
st38	13288 to	13302	n	-
st39	13398		n	-
st40	13910 to	13943	y	62E
st41	13984 to	14008	n	-
st42	14034 to	14085	n	-
st43	14104		n	-

Table 3: Combined medium-sized peaks and troughs

Code	Frame(s)		Up/ Down?	Nod No.
s1	3345 to	3371	y|y	16C
s2	6352 to	6365	n|n	-
s3	8428 to	8544	n|n	-
s4	9134 to	9150	n|n	-
s5	10438 to	10492	n|n	-
s6	13022 to	13084	y|y	57B
s7	13887 to	13943	y|y	62E
s8	14024 to	14085	n|n	-

KEY:		
sp	=	Medium-sized peaks
st	=	Medium-sized troughs
	=	Frames part of a peak-trough cluster
	=	Frames part of a peak-trough cluster
y	=	Yes (movement is a head nod)
n	=	No (movement is not a head nod)
	=	frame clusters aligned as a result of peak-trough clusters
	=	Data preceding the Case Study
No.	=	Nod number and coded type (see chapter 4)

Table 1: Intense and medium-sized head peaks

Group	Code	Frame(s)	Min.	Max.	Nod?	No.
	p	18 to 21	30		n	
	sp	22 to 92	29		n	
	p	86	30		n	N/A
	p	138 to 149	30	38	y	
	p	179 to 188	30	41	n	
cp1	sp1	238 to 243	29		n	-
cp2	sp2	304 to 381	29		n	-
cp3	sp3	863 to 871	29		y	4D
cp4	sp4	898 to 918	29		y	5C
cp5	sp5	1153 to 1154	29		y	6A
cp6	sp6	1272 to 1273	29		y	8A
cp7	sp7	1700	29		n	-
	p1	1721	30		n	
cp8	sp8	1790 to 1796	29		n	-
cp9	sp9	1949 to 1952	29		n	-
	p2	1966 to 1969	30	32	n	
cp10	sp10	2438 to 2444	29		y	11A
cp11	p3	2584 to 2599	30	31	y- nbc	-
cp12	sp11	2769 to 2776	29		n	-
cp13	sp12	2826	29		n	-
cp14	sp13	2974 to 2983	29		y	14B
	p4	2982	30			15B
cp15	sp14	3163 to 3171	29		n	-
	p5	3165 to 3166	29		n	
cp16	sp15	3270 to 3277	29		n	-
cp17	sp16	3362 to 3371	29		y	16C
	p6	3372 to 3396	30	32	n	
cp18	sp17	3637 to 3655	29		n	-
cp19	sp18	3891 to 3892	29		n	-
cp20	sp19	3948 to 4234	29		y	19B
	p7	3973 to 3980	30		y	17C
	p8	3998 to 4012	30	32	n	18A
	p9	4033 to 4060	30	31	n	
cp21	p10	4127 to 4132	30		n	-
	p11	4154 to 4167	30	32	n	
	p12	4192 to 4193	30		n	
	p13	4207 to 4208	30		n	
cp22	sp20	4283 to 4289	29		n	-
	p14	4284 to 4287	30	31	n	
	sp21	4297 to 4298	29		n	
cp23	p15	4326 to 4333	30	31	n	-
cp24	sp22	4674	29		y	22C
cp25	sp23	4837 to 4908	29		n	-
cp26	sp24	5735 to 5736	29		y	28B
	p16	5737 to 5753	30	32	y	29B
cp27	sp25	6352 to 6365	29		n	-
	p17	6354 to 6363	30	32	n	
cp28	sp26	6779 to 6812	29		n	-
cp29	p18	6813 to 6837	30	40	n	-
	sp27	6855 to 7026	29		n	
	p19	7000 to 7002	30		n	
	p20	7023 to 7025	30		n	
cp30	p21	7160 to 7162	30		n	-
	sp28	7163 to 7164	29		n	
cp31	sp29	7243 to 7249	29		n	-
	p22	7244 to 7245	30		n	
cp32	p23	7305 to 7310	30		n	-
	sp30	7311	29		n	
cp33	sp31	7558 to 7564	29		n	-
cp34	sp32	7705 to 7706	29		n	-
cp35	sp33	8377	29		n	-
cp36	sp34	8428 to 8434	29		n	-
	p24	8435 to 8437	31		n	
cp37	sp35	8521	29		n	-
cp38	sp36	8685 8730	29		n	-
	p25	8688 to 8693	30		n	
cp39	sp37	8778 to 8780	29		n	-
cp40	sp38	9023 to 9033	29		y	32C
	p26	9029 to 9030	30		y	33A
cp41	sp39	9134	29		n	-
cp42	sp40	9218 to 9221	29		n	-
	p27	9219 to 9220	30		n	
cp43	sp41	9249 to 9281	29		n	-
cp44	p28	9266 to 9269	30		n	-
	p29	9271 to 9278	30		n	
cp45	sp42	9318 to 9383	29		n	-
	sp30	9336 to 9338	30		n	
	p31	9351 to 9353	30		n	
cp46	sp43	10286 to 10302	29		n	-
cp47	sp44	10355 to 10382	29		n	-
	p32	10378	30	31	y	41C
cp48	p33	10435 to 10437	32		n	-
		10475 to 10493	30	32	y	42C
	sp45	10438	29		n	-
	sp46	10474 to 10492	29		n	-
cp49	sp47	10685 to 10746	29		n	-
cp50	p34	10686 to 10699	30		y	43C
cp51	sp48	11042 to 11048	29		y	44B
cp52	sp49	11370 to 11374	29		y	45C
	sp50	11392 to 11401	29		n	-
cp53	sp51	12007 to 12024	29		n	-
	p35	12021 to 12023	30		n	
	sp52	12218 to 12222	29		y	49A
cp54	sp53	12344 to 12366	29		y	50C
	p36	12351 to 12365	30		y	50C
cp55	sp54	12711 to 12750	29		n	-
	p37	12713 to 12725	30		y	54A
cp56	p38	12751 to 12775	30	31	y	55C
cp57	sp55	13049 to 13084	29		y	57B
	p39	13050 to 13062	30	32	y	57B
cp58	sp56	13110 to 13115	29		y	58C
	sp57	13131 to 13141	29		n	-
cp59	sp58	13204 to 13218	29		n	-
	p40	13214 to 13216	30	31	y	58C
cp60	sp59	13887 to 13903	29		y	62E
	p41	13897	30		y	62E
cp61	sp60	14024 to 14029	29		n	-
	p42	14028	30	32	n	
cp62	sp61	14135 to 14140	29		y	63A
cp63	sp62	14427 to 14451	29		n	-
cp64	sp63	15210 to 15230	29		y	70A

Table 2: Intense and medium-sized head troughs

Group	Code	Frame(s)	Min.	Max.	Nod?	No.
ct1	st1	466 to 469	21		y	2C
	t1	470 to 479	17	19	y	2C
ct2	st2	708 to 747	21		y	3A
ct3	st3	1310 to 1347	21		n	-
	t2	1318 to 1338	18	20	n	
ct4	st4	1454 to 1457	21		y	9B
	t3	1455 to 1456	20		y	9B
ct5	st5	1628 to 1632	21		y	10A
	st6	1658 to 1661	21		n	-
ct6	st7	2698 to 2709	21		n	-
	t4	2705	20		n	
ct7	st8	3345 to 3351	21		y	16C
	t5	3347 to 3348	20		n	-
	t6	3448 to 3451	20		n	-
	st9	3450 to 3455	21		y	16C
ct8	st10	3693 to 3768	21		y	17C
	t7	3720 to 3723	20		y	17C
ct9	t8	4377 to 4384	19	20	n	-
	st11	4379 to 4385	21		n	
ct10	st12	4426 to 4427	21		n	-
ct11	st13	4520 to 4529	21		y	21A
	t9	4526	20		n	-
ct12	st14	5371 to 5388	21		n	-
	t10	5383 to 5384	20		n	
ct13	st15	5901 to 5904	21		n	-
ct14	st16	5934 to 5946	21		y	29B
	t11	5947 to 6001	20		n	-
	st17	5993 to 5998	21		n	
ct15	st18	6376 to 6395	21		n	-
ct16	t12	6635 to 6645	18	20	n	-
	st19	6639	21		n	
ct17	st20	6724	21		n	-
ct18	st21	7369 to 7386	21		n	-
ct19	st22	7419 to 7458	21		n	-
	t13	7451 to 7464	18	20	n	
ct20	st23	8248 to 8264	21		n	-
	t14	8265	20		n	
ct21	st24	8451 to 8544	21		n	-
	t15	8546 to 8557	19	20	n	
ct22	st25	8864 to 8958	21		n	-
ct23	t16	8959	20		n	-
ct24	st26	9144 to 9150	21		n	-
	t17	9145 to 9156	17	20	y	32C
ct25	st27	10031 to 10116	21		y	40A
ct26	st28	10388 to 10411	21		y	41C
	t18	10389 to 10400	18	20	y	41C
ct27	st29	10449 to 10468	21		n	-
	t19	10450 to 10467	18	20	y	42C
ct28	st30	10604 to 10620	21		n	-
ct29	st31	10927 to 10936	21		y	43C
ct30	st32	11580	21		y	47B
ct31	st33	11881	21		n	-
ct32	t20	11941 to 11944	20		y-nbc	-
	st34	11939 to 11940	21		n	-
ct32	st35	11978 to 11979	21		n	-
ct33	st36	12043 to 12051	21		n	-
ct34	st37	13022	21		y	57B
	t21	13023 to 13046	16	20	y	57B
ct35	st38	13288 to 13302	21		n	-
ct36	st39	13398	21		n	-
ct37	st40	13910 to 13943	21		y	62E
	t22	13918 to 13944	19	20	y	62E
ct38	st41	13984 to 14008	21		n	-
	t23	13988 to 13992	20		n	-
ct39	st42	14034 to 14085	21		n	-
	t24	14035 to 14037	19	20	n	-
		14050 to 14061	18	20	n	-
		14086 to 14103	18	20	n	-
	st43	14104	21		n	-

Table 3: Combined peaks and troughs for intense and medium-sized nods

Code	Frame(s)	Min.	Max.	B/M Peaks?	B/M Troughs?	Nod No.
c1	2438 to 2709	20	29	p3|sp10	t4|st7	nbc|n|n|11A
c2	3345 to 3455	20	32	p6|sp16	t5/6|st8/9	n|n|16C|16C
c3	4326 to 4385	19	31	p15|n	t8|st11	n|n|n|n
c4	6352 to 6395	21	32	p17|sp25	t15|st18	n|n|n|n
c5	8428 to 8557	19	31	p24|sp34/5	t15|st24	n|n|n|n
c6	9134 to 9156	21	29	n|sp39	t17|st26	n|32C|n|n
c7	10355 to 10493	18	32	p32/3/42/3|sp4/5/6	t18|st28/9	41C/42C/43C/41C/42C|n|41C
c8	10435 to 10493	18	32	p35|sp51	t19|st36	n|n|n|n
c9	13022 to 13062	16	32	p39|sp55	n|st37	57B|57B|57B|57B
c10	13887 to 13944	19	32	p41|sp59	t21|st40	62E|n|62E|62E
c11	14024 to 14104	18	32	p42|sp60/1/2	t22|st42	n|n|63A|n

KEY:

p	=	Intense peaks
t	=	Intense troughs
sp	=	Medium-sized peaks
st	=	Medium-sized troughs
(yellow)	=	Frames part of a peak-trough cluster
(grey)	=	Frames part of a peak-trough cluster
y	=	Yes (movement is a head nod)
n	=	No (movement is not a head nod)
(hatched)	=	Frame clusters aligned as a result of peak-trough clusters
(blue outline)	=	Points where intense and medium-sized peaks and troughs cluster (within 25 frames)
(light blue)	=	Data preceding the Case Study
cp	=	Peaks across intense and/or medium-size movements (c = combined)
ct	=	Troughs across intense and/or medium-size movements (c = combined)
c	=	Combined peak and trough across the intense and/or medium-sized movements
nbc	=	Nod, but not a backchanneling nod
No.	=	Nod number and coded nod type (as ascribed in chapter 4)

Average = 25.33632
S.D. = 2.24729
2*S.D. = 4.49458
Exploring the range $21 \leq x \geq 29$
Combined Clusters: 25 frames $\leq\geq$ of each other (p/t)

Appendix 5.6: Combining the most intense and 'medium-sized' head peaks and troughs for S03MF.F, as seen across the case study excerpt (for a closer analysis of clusters of head movement).

Appendix 6.1: The analysis of S01FM.

The frequencies of specific spoken backchanneling forms / functions and (where relevant) concurrent backchanneling head nods

Section 1- S01MF.F

Spoken backchannel token / string	Function and Freq.	Without concurrent nods	With concurrent nods	A	B	C	D	E
Mmm	103	53	50	19	21	7	2	1
Yeah	45	8	37	12	9	11	4	1
	34	15	19	7	5	6	1	0
Uh huh	12	4	8	3	2	3	0	0
Yeah yeah	8	0	8	3	0	3	1	1
	4	0	4	0	0	2	1	1
Yep yeah	6	0	6	0	0	6	0	0
	1	1	0	0	0	0	0	0
Mmm mmm	5	1	4	0	2	0	0	2
Okay	3	3	0	0	0	0	0	0
	2	0	2	1	0	1	0	0
Yep	3	0	3	1	1	1	0	0
No	1	0	1	1	0	0	0	0
	1	1	0	0	0	0	0	0
	1	1	0	0	0	0	0	0
Right okay	2	1	1	0	0	1	0	0
Yeah mmm	2	2	0	0	0	0	0	0
Yeah okay	2	1	1	1	0	0	0	0
No no but you can	1	1	0	0	0	0	0	0
Oh no not at all no	1	1	0	0	0	0	0	0
Oh right	1	0	1	0	0	1	0	0
Oh that's interesting	1	1	0	0	0	0	0	0
Right yeah	1	0	1	0	0	1	0	0
Uh huh mmm	1	1	0	0	0	0	0	0
Yeah definitely yeah	1	1	0	0	0	0	0	0
Yeah err	1	0	1	0	0	0	0	1
Yeah I remember	1	1	0	0	0	0	0	0
Yeah no it will do I'm sure	1	0	1	0	0	1	0	0
Yeah small sample	1	1	0	0	0	0	0	0
Yeah that's a good way to think a link actually yeah	1	0	1	0	0	1	0	0
Yep that's really true	1	0	1	1	0	0	0	0
Yeah that's true yeah	1	1	0	0	0	0	0	0
Yeah you do	1	0	1	0	0	0	0	1
	250	**99**	**151**	**49**	**40**	**45**	**9**	**8**

Section 2- S01MF.M

Spoken backchannel token / string	Function and Freq.	Without concurrent nods	With concurrent nods	A	B	C	D	E
Yeah	14	4	10	8	1	1	0	0
	2	0	2	1	1	0	0	0
Mmm	8	5	3	3	0	0	0	0
Ok/okay	2	1	1	1	0	0	0	0
Alright	1	1	0	0	0	0	0	0
Definitely	1	0	1	0	1	0	0	0
I agree uh	1	1	0	0	0	0	0	0
Language	1	1	0	0	0	0	0	0
Mmm mmm	1	1	0	0	0	0	0	0
No	1	0	1	1	0	0	0	0
Uh	1	1	0	0	0	0	0	0
Well yeah	1	1	0	0	0	0	0	0
Yeah er	1	1	0	0	0	0	0	0
Yeah yeah	1	0	1	0	1	0	0	0
Yep	1	0	1	0	0	1	0	0
	37	**17**	**20**	**14**	**4**	**2**	**0**	**0**

KEY:

	Continuers
	Convergence Tokens
	Enagaged Response Tokens
	Information Receipt Tokens
A	Type A nod
B	Type B nod
C	Type C nod
D	Type D nod
E	Type E nod
BC	Spoken Backchannel

Closer analysis:

Section 3. Discourse functions of spoken backchannel forms

Function	Speaker	Frequency
Continuers (total = 177)	<$F>	163
	<$M>	14
Convergence Tokens (total = 86)	<$F>	68
	<$M>	18
Engaged Response Tokens (total = 12)	<$F>	9
	<$M>	2
Information Receipt Tokens (total = 12)	<$F>	10
	<$M>	3
		287

Section 4. Relationship between head nod type and discourse function

Type	Speaker	CON	CNV	ER	IR	Total
A	<$F>	29	18	1	1	49
	<$M>	5	8	0	1	14
B	<$F>	30	10	0	0	40
	<$M>	2	1	1	0	4
C	<$F>	18	23	2	2	45
	<$M>	0	2	0	0	2
D	<$F>	4	5	0	0	9
	<$M>	0	0	0	0	0
E	<$F>	5	3	0	0	8
	<$M>	0	0	0	0	0
		93	**70**	**4**	**4**	**171**

nbc = 15 for both <$1> and <$2>

Section 5. Frequencies of backchanneling head nods types

Type	Speaker	With spoken BC	Without spoken BC	Total
A	<$F>	49	16	65
	<$M>	14	47	61
B	<$F>	38	28	66
	<$M>	4	24	28
C	<$F>	44	12	56
	<$M>	2	8	10
D	<$F>	7	9	16
	<$M>	0	1	1
E	<$F>	6	2	8
	<$M>	0	0	0
Total		164	147	311

Section 6: Backchannels across turns

On occasion 1 nod is used across more than one verbal BC turn, details below:

Nod	Speaker	BC no	BC form
E	<$F>	Fd4	yeah
E	<$F>	Fd5	mmm
E	<$F>	Fd6	yeah err
B	<$F>	Ff8	yeah
B	<$F>	Ff9	yeah
E	<$F>	Fi8	yeah yeah
E	<$F>	Fi9	yeah you do
B	<$F>	Fj7	mmm
B	<$F>	Fj8	mmm
B	<$F>	Fk8	yeah
B	<$F>	Fk9	yeah
D	<$F>	Fs4	yeah
D	<$F>	Fs5	yeah

The frequencies of specific spoken backchanneling forms / functions and (where relevant) concurrent backchanneling head nods

Section 1- S02MM.1

Spoken backchannel token / string	Function and Freq.	Without concurrent nods	With concurrent nods	A	B	C	D	E
Yeah	9	5	4	2	0	1	1	0
	7	2	5	3	0	0	2	0
Sure	11	3	8	5	1	1	0	1
	1	0	1	1	0	0	0	0
	1	0	1	1	0	0	0	0
Right	6	2	4	0	1	3	0	0
	2	1	1	0	0	1	0	0
	2	0	2	0	0	2	0	0
Okay	5	1	4	3	0	0	0	1
	2	1	1	1	0	0	0	0
	1	0	1	0	1	0	0	0
Right okay	3	0	3	1	1	0	1	0
Good	2	1	1	0	0	0	0	1
Oh right	2	1	1	1	0	0	0	0
Yeah yeah	2	0	2	1	1	0	0	0
Yes	2	0	2	0	0	1	0	1
Absolutely absolutely	1	0	1	0	1	0	0	0
Excellent	1	0	1	1	0	0	0	0
Excellent. Yes.	1	0	1	1	0	0	0	0
Hm	1	0	1	0	0	0	1	0
Hmm	1	1	0	0	0	0	0	0
It's excellent	1	0	1	1	0	0	0	0
Mm	1	0	1	1	0	0	0	0
Okay. Right.	1	0	1	0	0	1	0	0
That's right	1	0	1	0	0	0	1	0
Uh huh	1	0	1	1	0	0	0	0
Uhm hm	1	0	1	1	0	0	0	0
Yep	1	0	1	0	0	1	0	0
	70	**18**	**52**	**25**	**6**	**11**	**6**	**4**

KEY:

Continuers (yellow)	
Convergence Tokens (cyan)	
Engaged Response Tokens (magenta)	
Information Receipt Tokens (red)	
A	Type A nod
B	Type B nod
C	Type C nod
D	Type D nod
E	Type E nod
BC	Spoken Backchannel

Section 2- S02MM.2

Spoken backchannel token / string	Function and Freq.	Without concurrent nods	With concurrent nods	A	B	C	D	E
Mm	243	40	203	47	83	13	12	48
Yeah	42	11	31	8	10	5	1	7
	14	4	10	1	4	2	0	3
Mm Mm	49	5	44	6	19	4	2	13
Uh hm	24	7	17	8	3	2	0	4
Sure	14	1	13	3	3	3	0	4
	2	1	1	1	0	0	0	0
	1	0	1	0	0	0	1	0
	1	0	1	0	1	0	0	0
Yes	16	2	14	4	4	5	0	1
Hm	9	2	7	2	5	0	0	0
Yeah Sure	3	1	2	1	0	0	0	1
	2	0	2	0	0	0	0	2
	1	0	1	0	0	0	0	1
Definitely	4	2	2	0	1	0	0	1
Mm mm mm	4	1	3	1	1	0	0	1
Sure Sure	3	0	3	1	1	0	0	1
Right	1	1	0	0	0	0	0	0
	1	0	1	0	1	0	0	0
That's right yes	2	1	1	0	0	0	0	1
Yeah Yeah	2	0	2	0	1	0	0	1
Yes Yeah	2	0	2	1	0	1	0	0
Definitely definitely	1	1	0	0	0	0	0	0
Exactly yeah er	1	1	0	0	0	0	0	0
Hm yeah	1	0	1	0	0	0	0	1
Interesting yeah	1	1	0	0	0	0	0	0
Mm <pause> sure	1	0	1	0	0	1	0	0
Mm. Sure.	1	0	1	0	1	0	0	0
Mm. That's right. Yes. Yeah.	1	0	1	0	0	1	0	0
Mm Yeah	1	1	0	0	0	0	0	0
No that's right	1	0	1	1	0	0	0	0
Oh exactly yeah	1	0	1	0	0	1	0	0
Oh okay. Yeah.	1	0	1	1	0	0	0	0
Oh wow	1	1	0	0	0	0	0	0
Okay	1	0	1	0	0	0	0	1
That's right. Er.	1	0	1	0	0	1	0	0
That's right	1	1	0	0	0	0	0	0
That's right. Yeah yeah.	1	0	1	0	0	1	0	0
Uh hm that's right yeah	1	0	1	0	0	1	0	0
Well yeah	1	1	0	0	0	0	0	0
yeah that's right yeah	1	1	0	0	0	0	0	0
Yeah that's that's interesting	1	1	0	0	0	0	0	0
Yeah yeah absolutely right	1	0	1	0	0	0	0	1
Yeah yeah yeah	1	0	1	0	1	0	0	0
Yeah yes	1	0	1	0	0	0	0	1
Yeah, that's important	1	1	0	0	0	0	0	0
Yep	1	0	1	0	0	0	0	1
Yes of course yes.	1	0	1	0	1	0	0	0
Yes yeah yeah yeah	1	0	1	0	0	0	0	1
Yes. Yeah. Mm.	1	1	0	0	0	0	0	0
Yes Yes Yes Yeah	1	0	1	0	0	0	0	1
	469	**90**	**379**	**86**	**140**	**41**	**16**	**96**

Closer analysis:

Section 3. Discourse functions of spoken backchannel forms

Function	Speaker	Frequency
Continuers (total = 358)	<$M.1>	13
	<$M.2>	345
Convergence Tokens (total = 126)	<$M.1>	33
	<$M.2>	93
Engaged Response Tokens (total = 34)	<$M.1>	10
	<$M.2>	24
Information Receipt Tokens (total = 21)	<$M.1>	14
	<$M.2>	7
		539

Section 4. Relationship between head nod type and discourse function

Type	Speaker	CON	CNV	ER	IR	Total
A	<$M.1>	6	12	4	3	25
	<$M.2>	65	18	2	1	86
B	<$M.1>	1	2	1	2	6
	<$M.2>	115	23	2	0	140
C	<$M.1>	0	5	2	4	11
	<$M.2>	21	15	4	1	41
D	<$M.1>	3	1	0	2	6
	<$M.2>	15	1	0	0	16
E	<$M.1>	0	3	1	0	4
	<$M.2>	70	19	4	3	96
		296	**99**	**20**	**16**	**431**

Section 5. Frequencies of backchanneling head nods types

nbc = 12 for <$M1> (0 for <$M2>)

Type	Speaker	With spoken BC	Without spoken BC	Total
A	<$M.1>	25	69	94
	<$M.2>	85	41	126
B	<$M.1>	5	13	18
	<$M.2>	95	34	129
C	<$M.1>	11	30	41
	<$M.2>	38	4	42
D	<$M.1>	6	1	7
	<$M.2>	11	0	11
E	<$M.1>	3	2	5
	<$M.2>	33	9	42
	Total	312	203	515

Appendix 6.2: The analysis of S02MM.

Section 6: Backchannels across turns

On occasion 1 nod is used across more than one verbal BC turn, details below:

Nod	Speaker	BC no	BC form
B	<$1>	M1e1	Right okay
B	<$1>	M1e2	Yeah yeah
E	<$1>	M1e6	Sure
E	<$1>	M1e7	Okay
B	<$2>	M2a1	Hm
B	<$2>	M2a2	Yeah
B	<$2>	M2a3	Yeah
B	<$2>	M2a4	Hm
B	<$2>	M2a8	Hm
B	<$2>	M2a9	Hm
B	<$2>	M2b1	Definitely
D	<$2>	M2b7	Mm
D	<$2>	M2b8	Mm Mm
E	<$2>	M2c3	Mm Mm
E	<$2>	M2c4	Definitely
E	<$2>	M2c5	Mm
E	<$2>	M2d2	Yep
E	<$2>	M2d3	Mm
C	<$2>	M2d4	Yeah
C	<$2>	M2d5	That's right. Er.
E	<$2>	M2d7	Mm
E	<$2>	M2d8	Mm
B	<$2>	M2e1	Mm Mm
B	<$2>	M2e2	Mm
B	<$2>	M2e3	Mm
D	<$2>	M2e8	Mm
D	<$2>	M2e9	Mm
E	<$2>	M2f1	Mm
E	<$2>	M2f2	Yes yeah yeah yeah
B	<$2>	M2f7	Mm
B	<$2>	M2f8	Mm
E	<$2>	M2h2	Uh hm
E	<$2>	M2h3	Mm Mm
E	<$2>	M2i8	Mm
E	<$2>	M2i9	Mm
E	<$2>	M2j4	Mm
E	<$2>	M2j5	Mm
B	<$2>	M2k9	Mm Mm
B	<$2>	M2l1	Mm
B	<$2>	M2l7	Mm
B	<$2>	M2l8	Mm
B	<$2>	M2l9	Mm
B	<$2>	M2m1	Mm
B	<$2>	M2m2	Mm
B	<$2>	M2n5	Mm
B	<$2>	M2n5	Yeah
E	<$2>	M2o5	Mm
E	<$2>	M2o6	Mm
E	<$2>	M2o7	Yeah
E	<$2>	M2o8	Mm
B	<$2>	M2p9	Mm
B	<$2>	M2q1	Yeah
E	<$2>	M2q2	Mm
E	<$2>	M2q3	Mm Mm
E	<$2>	M2q4	Mm
E	<$2>	M2q5	Mm
E	<$2>	M2q6	Mm
B	<$2>	M2r2	Mm
B	<$2>	M2r3	Mm
C	<$2>	M2r7	Mm
C	<$2>	M2r8	Yeah
B	<$2>	M2r9	Mm
B	<$2>	M2s1	Mm
B	<$2>	M2u4	Mm
B	<$2>	M2u5	Mm
B	<$2>	M2u6	Mm. Sure.
B	<$2>	M2u7	Mm
B	<$2>	M2u8	Mm
C	<$2>	M2u9	Yeah
C	<$2>	M2v1	That's right. Yeah yeah.
B	<$2>	M2w6	Yeah
B	<$2>	M2w7	Yeah
B	<$2>	M2w8	Mm
E	<$2>	M2x3	Mm Mm
E	<$2>	M2x4	Yeah sure
E	<$2>	M2x5	Mm Mm
E	<$2>	M2x6	Mm
B	<$2>	M2y1	Mm
B	<$2>	M2y2	Mm
B	<$2>	M2y3	Sure
B	<$2>	M2z1	Uh hm
B	<$2>	M2z2	Yeah
E	<$2>	M2z6	Yeah
E	<$2>	M2z7	Mm Mm
E	<$2>	M2a.1	Mm
E	<$2>	M2a.2	Mm
B	<$2>	M2a.3	Mm
B	<$2>	M2a.4	Mm
E	<$2>	M2a.5	Yes
E	<$2>	M2a.6	Mm
E	<$2>	M2a.8	Mm mm
E	<$2>	M2a.9	Yeah sure
E	<$2>	M2b.1	Mm Mm Mm
E	<$2>	M2b.2	Mm
B	<$2>	M2b.6	Mm Mm
B	<$2>	M2b.7	Mm
B	<$2>	M2e.6	Mm
B	<$2>	M2e.7	Mm
E	<$2>	M2f.7	Mm
E	<$2>	M2f.8	Mm
E	<$2>	M2g.3	Yeah
E	<$2>	M2g.4	Mm
E	<$2>	M2g.5	Mm
B	<$2>	M2g.9	Mm
B	<$2>	M2h.1	Mm
B	<$2>	M2i.8	Mm
B	<$2>	M2i.9	Yeah
E	<$2>	M2j.2	Yeah
E	<$2>	M2j.3	Mm Mm
D	<$2>	M2j.5	Mm
D	<$2>	M2j.6	Mm
D	<$2>	M2j.7	Mm
E	<$2>	M2k.6	Mm
E	<$2>	M2k.7	Mm
E	<$2>	M2k.8	Mm
E	<$2>	M2k.9	Mm
E	<$2>	M2l.1	Mm
B	<$2>	M2l.3	Mm
B	<$2>	M2l.4	Mm Mm

Nod	Speaker	BC no	BC form
E	<$2>	M2l.9	Sure
E	<$2>	M2m.1	Yeah
B	<$2>	M2n.1	Mm
B	<$2>	M2n.2	Mm Mm
B	<$2>	M2n.3	Mm
E	<$2>	M2o.2	Yeah. Sure.
E	<$2>	M2o.3	Okay
E	<$2>	M2o.6	Sure
E	<$2>	M2o.7	Yes Yes Yes Yeah
E	<$2>	M2o.8	Mm
E	<$2>	M2o.9	Mm
E	<$2>	M2p.3	Mm
E	<$2>	M2p.4	Mm
E	<$2>	M2p.5	Yeah
B	<$2>	M2p.9	Mm
B	<$2>	M2q.1	Mm
E	<$2>	M2r.6	Mm
E	<$2>	M2r.7	Yeah
E	<$2>	M2r.8	Uh hm
B	<$2>	M2r.9	Mm
B	<$2>	M2s.1	Yeah
B	<$2>	M2s.2	Mm
E	<$2>	M2u.2	Sure
E	<$2>	M2u.3	Mm Mm
E	<$2>	M2u.4	Yeah yeah absolutely right
B	<$2>	M2u.7	Yeah yeah yeah
B	<$2>	M2u.8	Mm Mm
E	<$2>	M2v.1	Mm
E	<$2>	M2v.2	Yeah yes
E	<$2>	M2v.4	Mm
E	<$2>	M2v.5	Mm
E	<$2>	M2v.6	Mm
E	<$2>	M2v.7	Mm
E	<$2>	M2v.8	Yeah Yeah
E	<$2>	M2v.9	That's right yes
E	<$2>	M2w.1	Uh hm
E	<$2>	M2w.2	Mm
E	<$2>	M2w.3	Mm
E	<$2>	M2w.4	Mm
E	<$2>	M2x.6	Yeah
E	<$2>	M2x.7	Sure Sure
E	<$2>	M2x.8	Sure
B	<$2>	M2y.1	Mm
B	<$2>	M2y.2	Mm
B	<$2>	M2y.6	Mm
B	<$2>	M2y.7	Mm
E	<$2>	M2z.4	Yeah sure
E	<$2>	M2z.5	Mm

The frequencies of specific spoken backchanneling forms / functions and (where relevant) concurrent backchanneling head nods

Section 1- S03MF.M

Spoken backchannel token / string	Function and Freq.	Without concurrent nods	With concurrent nods	Nod type A	B	C	D	E
Yeah	40	15	25	5	13	3	0	4
	13	6	7	3	2	2	0	0
Right	23	11	12	3	3	5	0	1
	5	2	3	1	1	1	0	0
	1	1	0	0	0	0	0	0
Yeah yeah	20	4	16	5	7	1	2	1
	1	0	1	0	0	1	0	0
Okay	6	1	5	2	0	3	0	0
	2	1	1	1	0	0	0	0
Mm	3	2	1	1	0	0	0	0
Oh yeah	1	0	1	0	0	0	0	1
	1	1	0	0	0	0	0	0
Right yeah	1	0	1	1	0	0	0	0
	1	0	1	0	0	1	0	0
Right yeah yeah	2	2	0	0	0	0	0	0
Uh-huh	2	0	2	0	0	1	1	0
Yeah yeah yeah	2	0	2	1	1	0	0	0
Ah okay	1	1	0	0	0	0	0	0
Ah right. Okay.	1	1	0	0	0	0	0	0
Aha	1	0	1	1	0	0	0	0
Er	1	1	0	0	0	0	0	0
Erm	1	1	0	0	0	0	0	0
Erm yeah	1	1	0	0	0	0	0	0
Erm yeah yeah okay	1	0	1	0	1	0	0	0
Oh does it	1	1	0	0	0	0	0	0
Oh god	1	1	0	0	0	0	0	0
Oh I see right	1	1	0	0	0	0	0	0
Oh really	1	1	0	0	0	0	0	0
Oh really? Oh right	1	1	0	0	0	0	0	0
Oh right oh okay	1	1	0	0	0	0	0	0
Oh right yeah	1	0	1	0	0	0	0	1
Oh wow. Right.	1	1	0	0	0	0	0	0
Okay brilliant	1	0	1	0	0	1	0	0
Okay yeah brilliant	1	0	1	1	0	0	0	0
Right yeah yeah yeah	1	0	1	0	1	0	0	0
Right. Ah right.	1	1	0	0	0	0	0	0
Right. Oh right yeah.	1	0	1	1	0	0	0	0
Right. Okay.	1	0	1	1	0	0	0	0
Sure yeah yeah	1	0	1	1	0	0	0	0
That's right	1	1	0	0	0	0	0	0
Well yeah yeah	1	0	1	0	1	0	0	0
Yeah oh god that yeah	1	1	0	0	0	0	0	0
Yeah okay	1	0	1	1	0	0	0	0
Yeah right good no it looks really good yeah	1	0	1	0	1	0	0	0
Yeah right yeah	1	1	0	0	0	0	0	0
Yeah something like that	1	1	0	0	0	0	0	0
Yeah that's right	1	0	1	0	0	1	0	0
Yeah that's right yeah	1	0	1	0	0	0	1	0
Yeah yeah er	1	0	1	1	0	0	0	0
Yeah yeah of course it is yeah	1	0	1	1	0	0	0	0
Yeah yeah yeah yeah	1	0	1	0	1	0	0	0
Yeah yeah yeah yeah that's right	1	0	1	0	0	0	0	1
Yeah. Right.	1	0	1	0	1	0	0	0
	160	63	97	31	33	20	4	9

Section 2- S03MF.F

Spoken backchannel token / string	Function and Freq.	Without concurrent nods	With concurrent nods	Nod type A	B	C	D	E
Yeah	44	6	38	22	2	11	1	2
	30	2	28	14	3	11	0	0
	1	0	1	1	0	0	0	0
Okay	9	2	7	6	0	1	0	0
	5	1	4	1	1	2	0	0
Right	11	4	7	4	0	3	0	0
	3	0	3	2	0	1	0	0
Erm	5	5	0	0	0	0	0	0
Mm	2	1	1	1	0	0	0	0
No	2	1	1	0	0	1	0	0
Yeah erm	1	1	0	0	0	0	0	0
	1	0	1	1	0	0	0	0
Is it?	1	1	0	0	0	0	0	0
No okay	1	0	1	1	0	0	0	0
Oh right	1	0	1	0	0	1	0	0
Oh right I see	1	0	1	0	0	1	0	0
Oh right okay	1	0	1	1	0	0	0	0
Oh that should be okay	1	0	1	0	1	0	0	0
Quite interesting	1	1	0	0	0	0	0	0
Right yeah	1	0	1	1	0	0	0	0
Right yeah I do that yeah	1	0	1	0	1	0	0	0
That would be ideal yeah	1	0	1	1	0	0	0	0
Yeah <pause> erm	1	0	1	1	0	0	0	0
Yeah er	1	0	1	0	0	1	0	0
Yeah I think so	1	0	1	0	0	1	0	0
Yeah it would	1	0	1	0	0	1	0	0
Yeah twelve I think	1	1	0	0	0	0	0	0
Yeah uh-huh	1	0	1	0	0	1	0	0
Yeah yeah	1	0	1	0	0	1	0	0
Yes	1	0	1	0	0	1	0	0
	132	26	106	57	8	38	1	2

Appendix 6.3: The analysis of S03MF.

Closer analysis:

Section 3. Discourse functions of spoken backchannel forms

Function	Speaker	Frequency
Continuers (total = 103)	<$M>	51
	<$F>	52
Convergence Tokens (total = 111)	<$M>	50
	<$F>	61
Engaged Response Tokens (total = 25)	<$M>	17
	<$F>	8
Information Receipt Tokens (total = 53)	<$M>	42
	<$F>	11
		292

Section 4. Relationship between head nod type and discourse function

Type	Speaker	CON	CNV	ER	IR	Total
A	<$M>	7	13	2	9	31
	<$F>	23	27	2	5	57
B	<$M>	13	15	1	4	33
	<$F>	2	3	2	1	8
C	<$M>	5	5	0	10	20
	<$F>	11	22	1	4	38
D	<$M>	1	2	1	0	4
	<$F>	1	0	0	0	1
E	<$M>	4	1	3	1	9
	<$F>	2	0	0	0	2
		69	**88**	**12**	**34**	**203**

Section 5. Frequencies of backchanneling head nods types

nbc = 8 for <$1> and 18 for <$2>

Type	Speaker	With spoken BC	Without spoken BC	Total
A	<$M>	31	35	66
	<$F>	57	146	203
B	<$M>	23	17	40
	<$F>	8	22	30
C	<$M>	20	12	32
	<$F>	38	36	74
D	<$M>	4	4	8
	<$F>	1	0	1
E	<$M>	5	3	8
	<$F>	2	1	3
	Total	**189**	**276**	**465**

Section 6: Backchannels across turns

On occasion 1 nod is used across more than one verbal BC turn, details below:

Nod	Speaker	BC no	BC form
B	<$M>	1c4	Yeah yeah
B	<$M>	1c5	Yeah yeah
B	<$M>	1c6	Yeah yeah
B	<$M>	1e 2	Yeah
B	<$M>	1e 3	Right
B	<$M>	1f7	Yeah
B	<$M>	1f8	Yeah
B	<$M>	1g9	Right
B	<$M>	1h1	Yeah
B	<$M>	1h2	Yeah. Right.
E	<$M>	1h3	Yeah
E	<$M>	1h4	Yeah yeah yeah yeah that's right
B	<$M>	1h5	Yeah
B	<$M>	1h6	Yeah
B	<$M>	1h7	Yeah yeah
B	<$M>	1h9	Yeah
B	<$M>	1i1	Yeah
B	<$M>	1i2	Yeah yeah
E	<$M>	1k9	Right
E	<$M>	1l1	Yeah yeah
E	<$M>	1n5	Oh right yeah
E	<$M>	1n6	Yeah
E	<$M>	1n4	Yeah

KEY:

	Continuers
	Convergence Tokens
	Engaged Response Tokens
	Information Receipt Tokens
A	Type A nod
B	Type B nod
C	Type C nod
D	Type D nod
E	Type E nod
BC	Spoken Backchannel

Appendix 6.4: The analysis of S04MM.

The frequencies of specific spoken backchanneling forms / functions and (where relevant) concurrent backchanneling head nods

Section 1- S04MM.1

Spoken backchannel token / string	Function and Freq.	Without concurrent nods	With concurrent nods	Nod type A	B	C	D	E
Mm	104	40	64	29	11	11	8	5
Yeah	42	12	30	9	8	1	7	5
	41	15	26	5	4	5	7	5
Yes	25	9	16	7	3	1	4	1
Okay	7	1	6	2	0	2	2	0
	4	3	1	0	0	0	1	0
Mhm	7	3	4	0	1	3	0	0
Yeah okay	3	2	1	1	0	0	0	0
	2	0	2	0	0	0	1	1
Yeah yeah	4	1	3	1	1	1	0	0
	1	0	1	0	1	0	0	0
No	2	2	0	0	0	0	0	0
	1	1	0	0	0	0	0	0
Yeah mm	3	3	0	0	0	0	0	0
Yes yeah	2	0	2	0	1	0	1	0
	1	1	0	0	0	0	0	0
Okay yeah	2	0	2	1	0	0	1	0
Right	2	1	1	0	1	0	0	0
Hm	1	0	1	1	0	0	0	0
Interesting	1	1	0	0	0	0	0	0
Interesting interesting	1	0	1	1	0	0	0	0
Interesting isn't it	1	1	0	0	0	0	0	0
Mm mm	1	1	0	0	0	0	0	0
Mm yes there are quite a few	1	0	1	0	1	0	0	0
Mm. Interesting isn't it.	1	0	1	0	0	1	0	0
Of sixty	1	1	0	0	0	0	0	0
Oh is it	1	1	0	0	0	0	0	0
Oh okay	1	0	1	1	0	0	0	0
Oh that sort of thing	1	1	0	0	0	0	0	0
Oh yeah	1	0	1	0	0	1	0	0
Ok good	1	0	1	1	0	0	0	0
Okay I've got you	1	0	1	0	0	0	1	0
Okay. Yeah. Mm.	1	0	1	0	0	0	1	0
Sorry. Yeah.	1	0	1	0	0	1	0	0
Sure	1	0	1	1	0	0	0	0
That sort of idea. Yeah.	1	0	1	1	0	0	0	0
Yeah no go on	1	1	0	0	0	0	0	0
Yeah okay mm	1	0	1	0	1	0	0	0
Yeah right	1	0	1	0	0	0	1	0
Yeah that one yeah	1	1	0	0	0	0	0	0
Yeah they do	1	0	1	0	1	0	0	0
Yeah we have. Mm.	1	0	1	0	0	0	0	1
Yeah. Interesting	1	0	1	1	0	0	0	0
Yeah. Mm okay.	1	0	1	0	0	0	1	0
Yes I know	1	1	0	0	0	0	0	0
Yes I've got you	1	0	1	0	1	0	0	0
Yes that type of thing	1	0	1	0	1	0	0	0
Yes. Yeah yeah yeah yeah.	1	0	1	0	0	0	1	0
	283	**103**	**180**	**62**	**36**	**27**	**37**	**18**

KEY:

Continuers	(yellow)
Convergence Tokens	(cyan)
Engaged Response Tokens	(magenta)
Information Receipt Tokens	(red)
A	Type A nod
B	Type B nod
C	Type C nod
D	Type D nod
E	Type E nod
BC	Spoken Backchannel

Section 2- S04MM.2

Spoken backchannel token / string	Function and Freq.	Without concurrent nods	With concurrent nods	Nod type A	B	C	D	E
Yeah	76	25	51	21	22	6	2	0
	73	31	42	23	11	5	1	2
Mm	48	13	35	19	16	0	0	0
Right	8	3	5	1	0	4	0	0
	8	4	4	4	0	0	0	0
Yeah yeah	9	3	6	3	2	0	1	0
	1	0	1	0	1	0	0	0
Mhm	6	0	6	5	1	0	0	0
	1	1	0	0	0	0	0	0
Mm yeah	4	0	4	3	0	1	0	0
Okay	4	4	0	0	0	0	0	0
No	2	0	2	2	0	0	0	0
Yes	2	0	2	2	0	0	0	0
Empty yes	1	0	1	0	1	0	0	0
Erm	1	1	0	0	0	0	0	0
Is it? Oh	1	1	0	0	0	0	0	0
Oh I know yeah	1	1	0	0	0	0	0	0
Right yeah	1	1	0	0	0	0	0	0
Yeah erm	1	1	0	0	0	0	0	0
Yeah yeah absolutely	1	0	1	0	1	0	0	0
Yes. Yeah.	1	1	0	0	0	0	0	0
	250	**90**	**160**	**83**	**55**	**16**	**4**	**2**

Closer analysis:

Section 3. Discourse functions of spoken backchannel forms

Function	Speaker	Frequency
Continuers (total = 289)	<$1>	157
	<$2>	132
Convergence Tokens (total = 191)	<$1>	89
	<$2>	102
Engaged Response Tokens (total = 28)	<$1>	24
	<$2>	4
Information Receipt Tokens (total = 25)	<$1>	13
	<$2>	12
		533

Section 4. Relationship between head nod type and discourse function

Type	Speaker	CON	CNV	ER	IR	Total
A	<$1>	39	14	6	3	62
	<$2>	45	37	0	1	83
B	<$1>	21	10	5	0	36
	<$2>	40	13	2	0	55
C	<$1>	15	8	2	2	27
	<$2>	6	6	0	4	16
D	<$1>	15	16	2	4	37
	<$2>	2	2	0	0	4
E	<$1>	10	7	1	0	18
	<$2>	0	2	0	0	2
		193	**115**	**18**	**14**	**340**

Section 5. Frequencies of backchanneling head nods types

Type	Speaker	With spoken BC	Without spoken BC	Total
A	<$1>	62	20	82
	<$2>	83	61	144
B	<$1>	22	18	40
	<$2>	41	34	75
C	<$1>	27	10	37
	<$2>	16	10	26
D	<$1>	21	12	33
	<$2>	3	0	3
E	<$1>	7	2	9
	<$2>	1	3	4
Total		283	170	453

nbc = 41 for <$1> and 3 for <$2>

Appendix 6.4: The analysis of S04MM.

Section 6: Backchannels across turns

On occasion 1 nod is used across more than one verbal BC turn, details below:

Nod	Speaker	BC no	BC form
D	<$1>	M1a9	Yeah
D	<$1>	M1b1	Yeah
E	<$1>	M1d6	Yeah we have. Mm.
E	<$1>	M1d7	Mm
B	<$1>	M1e6	Mm
B	<$1>	M1e7	Mm
D	<$1>	M1f4	Mm
D	<$1>	M1f5	Mm
D	<$1>	M1f6	Mm
B	<$1>	M1f8	Yeah
B	<$1>	M1f9	Yeah
B	<$1>	M1g1	Yeah
D	<$1>	M1g8	Okay
D	<$1>	M1g9	Mm
D	<$1>	M1h1	Okay
D	<$1>	M1h2	Mm
B	<$1>	M1i2	Yeah they do
B	<$1>	M1i3	Yeah yeah
E	<$1>	M1k5	Yes okay
E	<$1>	M1k6	Yeah
E	<$1>	M1k7	Yeah
E	<$1>	M1k8	Yeah
D	<$1>	M1l5	Okay
D	<$1>	M1l6	Mm
B	<$1>	M1l8	Yeah yeah
B	<$1>	M1l9	Yeah
D	<$1>	M1n4	Yeah
D	<$1>	M1n5	Yeah
E	<$1>	M1n6	Mm
E	<$1>	M1n7	Yes
E	<$1>	M1n8	Yeah
B	<$1>	M1o4	Mm
B	<$1>	M1o5	Yeah
B	<$1>	M1o6	Yeah
B	<$1>	M1o7	Yeah
D	<$1>	M1q8	Yes yeah
D	<$1>	M1q9	Yeah
E	<$1>	M1r2	Yeah
E	<$1>	M1r3	Yeah
E	<$1>	M1r4	Mm
E	<$1>	M1r5	Mm
D	<$1>	M1s7	Yeah
D	<$1>	M1s8	Mm
E	<$1>	M1u6	Yeah
E	<$1>	M1u7	Yeah
B	<$1>	M1v6	Yes
B	<$1>	M1v7	Yes
B	<$1>	M1v8	Mhm
B	<$1>	M1v9	Mm
B	<$1>	M1w1	Mm
D	<$1>	M1y2	Yeah
D	<$1>	M1y3	Yeah
D	<$1>	M1y4	Yeah
D	<$1>	M1y6	Yes
D	<$1>	M1y7	Yeah
E	<$1>	M1z1	Yeah
E	<$1>	M1z2	Yeah
B	<$1>	M1z4	Yeah
B	<$1>	M1z5	Mm
B	<$1>	M1z6	Mm
B	<$1>	M1c.5	Yeah
B	<$1>	M1c.6	Mm
D	<$1>	M1c.9	Okay I've got you
D	<$1>	M1d.1	Yes
D	<$1>	M1d.2	Yeah
D	<$1>	M1d.3	Yeah right
D	<$1>	M1e.7	Yes
D	<$1>	M1e.8	Yeah

Nod	Speaker	BC no	BC form
B	<$2>	M2c5	Yeah
B	<$2>	M2c6	Yeah
E	<$2>	M2c7	Yeah
E	<$2>	M2c8	Yeah
B	<$2>	M2d2	Empty yes
B	<$2>	M2d3	Yeah
B	<$2>	M2d9	Yeah
B	<$2>	M2e1	Yeah
D	<$2>	M2g5	Yeah yeah
D	<$2>	M2g6	Yeah
B	<$2>	M2g7	Yeah
B	<$2>	M2g8	Yeah
B	<$2>	M2k1	Mm
B	<$2>	M2k2	Yeah
B	<$2>	M2m1	Mm
B	<$2>	M2m2	Yeah
B	<$2>	M2u5	Yeah
B	<$2>	M2u6	Yeah
B	<$2>	M2u7	Yeah
B	<$2>	M2w1	Mm
B	<$2>	M2w2	Mm
B	<$2>	M2w3	Mm
B	<$2>	M2w9	Mm
B	<$2>	M2x1	Yeah
B	<$2>	M2y3	Mm
B	<$2>	M2y4	Yeah
B	<$2>	M2z2	Yeah
B	<$2>	M2z3	Yeah
B	<$2>	M2a.7	Yeah
B	<$2>	M2a.8	Yeah

Appendix 6.5: The analysis of S05MM.

The frequencies of specific spoken backchanneling forms / functions and (where relevant) concurrent backchanneling head nods

Section 1- S05MM.1

Spoken backchannel token / string	Function and Freq.	Without concurrent nods	With concurrent nods	A	B	C	D	E
Mm	179	39	140	85	50	4	0	1
Yes	92	24	68	36	14	12	2	4
Yeah	15	6	9	3	3	3	0	0
	12	4	8	4	2	2	0	0
Okay	5	1	4	2	1	1	0	0
	1	0	1	0	0	1	0	0
That's right	3	1	2	0	1	1	0	0
	2	0	2	0	0	0	1	1
Right	2	2	0	0	0	0	0	0
	2	1	1	1	0	0	0	0
No	3	1	2	1	1	0	0	0
Mm mm (or mm mmm)	3	0	3	3	0	0	0	0
Sure	2	0	2	1	0	1	0	0
Yeah that's right	2	1	1	1	0	0	0	0
Yes mm	2	0	2	1	1	0	0	0
Yes yes	2	0	2	1	1	0	0	0
Absolutely yeah	1	0	1	1	0	0	0	0
Ah yes <laughs>	1	1	0	0	0	0	0	0
Aha	1	1	0	0	0	0	0	0
Er yes	1	0	1	0	0	1	0	0
Mm that's right. Yeah.	1	0	1	0	1	0	0	0
Mm yes that's true	1	0	1	0	0	0	0	1
No no	1	0	1	1	0	0	0	0
No that's right	1	0	1	0	1	0	0	0
No that's right no	1	0	1	1	0	0	0	0
Oh I see what you mean yeah	1	0	1	1	0	0	0	0
That's true	1	0	1	0	0	0	0	1
That's true yeah	1	0	1	1	0	0	0	0
Well that's right exactly	1	0	1	0	1	0	0	0
Yeah no	1	0	1	0	0	1	0	0
Yes yes yes	1	0	1	0	1	0	0	0
	342	**82**	**260**	**144**	**78**	**27**	**3**	**8**

Section 2- S05MM.2

Spoken backchannel token / string	Function and Freq.	Without concurrent nods	With concurrent nods	A	B	C	D	E
Mm	89	21	68	56	2	10	0	0
Yeah	33	7	26	15	2	9	0	0
	8	4	4	1	2	1	0	0
Yeah absolutely	3	0	3	2	0	1	0	0
Right	2	1	1	1	0	0	0	0
Yeah yeah	2	0	2	2	0	0	0	0
Definitely	1	1	0	0	0	0	0	0
Mm yeah absolutely	1	0	1	1	0	0	0	0
Okay	1	0	1	0	0	1	0	0
Right okay	1	0	1	0	0	1	0	0
Well yeah absolutely yeah	1	0	1	0	0	1	0	0
Yeah definitely	1	0	1	0	0	1	0	0
Yeah well yeah	1	0	1	0	0	0	0	1
Yes	1	1	0	0	0	0	0	0
	145	**35**	**110**	**78**	**6**	**25**	**0**	**1**

Closer analysis:

Section 3. Discourse functions of spoken backchannel forms

Function	Speaker	Frequency
Continuers (total = 294)	<$1>	197
	<$2>	97
Convergence Tokens (total = 158)	<$1>	120
	<$2>	38
Engaged Response Tokens (total = 22)	<$1>	16
	<$2>	6
Information Receipt Tokens (total = 13)	<$1>	9
	<$2>	4
		487

Section 4. Relationship between head nod type and discourse function

Type	Speaker	CON	CNV	ER	IR	Total
A	<$1>	91	45	5	3	144
	<$2>	57	17	3	1	78
B	<$1>	53	20	4	1	78
	<$2>	4	2	0	0	6
C	<$1>	7	18	1	1	27
	<$2>	11	10	2	2	25
D	<$1>	0	2	0	1	3
	<$2>	0	0	0	0	0
E	<$1>	1	4	2	1	8
	<$2>	0	1	0	0	1
		224	**119**	**17**	**10**	**370**

Section 5. Frequencies of backchanneling head nods types

Type	Speaker	With spoken BC	Without spoken BC	Total
A	<$1>	144	32	176
	<$2>	78	59	137
B	<$1>	55	17	72
	<$2>	6	5	11
C	<$1>	27	4	31
	<$2>	25	12	37
D	<$1>	2	0	2
	<$2>	0	1	1
E	<$1>	5	0	5
	<$2>	1	0	1
	Total	343	130	473

nbc = 11 for <$1> and 11 for <$2>

Section 6: Backchannels across turns

On occasion 1 nod is used across more than one verbal BC turn, details below:

Nod	Speaker	BC no	BC form
B	<$1>	M1a9	Yes yes yes
B	<$1>	M1b1	Yeah
B	<$1>	M1c1	Yes
B	<$1>	M1c2	Yes
B	<$1>	M1e1	Mm
B	<$1>	M1e2	Mm
D	<$1>	M1f4	Yes
D	<$1>	M1f5	Yes
B	<$1>	M1g3	Mm
B	<$1>	M1g4	Yes
E	<$1>	M1g5	That's right
E	<$1>	M1g6	Yes
B	<$1>	M1l3	Mm
B	<$1>	M1l4	Mm that's right. Yeah.
B	<$1>	M1l8	Mm
B	<$1>	M1l9	Mm
B	<$1>	M1n8	Mm
B	<$1>	M1n9	Mm
B	<$1>	M1p1	No
B	<$1>	M1p2	Mm
B	<$1>	M1q3	Mm
B	<$1>	M1q4	Mm
B	<$1>	M1q6	Mm
B	<$1>	M1q7	Mm
B	<$1>	M1r8	Yes
B	<$1>	M1r9	Yeah
B	<$1>	M1s1	Yeah
B	<$1>	M1t2	Mm
B	<$1>	M1t3	Mm
E	<$1>	M1z9	Yes
E	<$1>	M1a.1	That's true
E	<$1>	M1a.2	Yes
B	<$1>	M1a.7	Yes
B	<$1>	M1a.8	Mm
B	<$1>	M1c.8	Mm
B	<$1>	M1c.9	Mm
B	<$1>	M1d.4	Mm
B	<$1>	M1d.5	Mm
B	<$1>	M1g.9	Mm
B	<$1>	M1h.1	Yes
B	<$1>	M1h.4	Yes
B	<$1>	M1h.5	Mm
B	<$1>	M1k.3	Mm
B	<$1>	M1k.4	Mm
B	<$1>	M1k.6	That's right
B	<$1>	M1k.7	Yes
B	<$1>	M1k.8	Yes
B	<$1>	M1k.9	Well that's right exactly
B	<$1>	M1l.1	Yeah
B	<$1>	M1l.5	Mm
B	<$1>	M1l.6	Mm

KEY:

	Continuers
	Convergence Tokens
	Engaged Response Tokens
	Information Receipt Tokens
A	Type A nod
B	Type B nod
C	Type C nod
D	Type D nod
E	Type E nod
BC	Spoken Backchannel

Appendix 6.6: The analysis of S06FF.

The frequencies of specific spoken backchanneling forms / functions and (where relevant) concurrent backchanneling head nods

Section 1- S06FF.1

Spoken backchannel token / string	Function and Freq.	Without concurrent nods	With concurrent nods	A	B	C	D	E
Right	22	3	19	8	1	6	2	2
	9	2	7	4	0	3	0	0
Yes	25	1	24	6	1	4	5	8
	1	0	1	0	1	0	0	0
	1	0	1	0	0	0	1	0
Yeah	15	1	14	3	1	5	2	3
	6	3	3	2	0	1	0	0
Mhm	5	2	3	2	0	0	1	0
Mm	4	1	3	3	0	0	0	0
Yes Yeah	1	0	1	0	0	0	1	0
	1	0	1	0	0	0	0	1
Absolutely yes	1	0	1	0	1	0	0	0
Absolutely yes yes	1	0	1	0	1	0	0	0
Ah right	1	1	0	0	0	0	0	0
I know	1	0	1	0	0	0	0	1
I see what you mean	1	0	1	0	0	1	0	0
It does yeah	1	0	1	1	0	0	0	0
Okay	1	1	0	0	0	0	0	0
That's okay	1	1	0	0	0	0	0	0
That's right yeah	1	0	1	0	0	1	0	0
Yeah I know. Yes.	1	0	1	0	1	0	0	0
Yeah yeah	1	0	1	0	1	0	0	0
Yes absolutely right yes	1	0	1	0	0	0	1	0
Yes I know	1	0	1	0	1	0	0	0
Yes mhm	1	0	1	1	0	0	0	0
Yes that's right	1	0	1	0	1	0	0	0
	105	**16**	**89**	**30**	**10**	**21**	**13**	**15**

Section 2- S06FF.2

Spoken backchannel token / string	Function and Freq.	Without concurrent nods	With concurrent nods	A	B	C	D	E
Mhm	83	17	66	8	35	4	7	12
	1	0	1	0	0	0	1	0
Yeah	51	17	34	7	14	4	2	7
	5	1	4	2	1	1	0	0
Okay	18	8	10	4	2	2	1	1
	2	1	1	0	0	1	0	0
Mm	9	1	8	2	3	1	1	1
Right	6	2	4	0	0	3	1	0
Yeah true	1	0	1	0	1	0	0	0
	1	1	0	0	0	0	0	0
Definitely	1	1	0	0	0	0	0	0
Erm	1	1	0	0	0	0	0	0
Mhm. That's fine.	1	0	1	0	0	0	0	1
Right yeah	1	1	0	0	0	0	0	0
Simplify	1	0	1	0	1	0	0	0
True	1	0	1	0	1	0	0	0
Yeah erm	1	1	0	0	0	0	0	0
Yeah I know	1	1	0	0	0	0	0	0
Yeah okay	1	0	1	0	1	0	0	0
Yes	1	1	0	0	0	0	0	0
	187	**54**	**133**	**23**	**59**	**16**	**13**	**22**

Closer analysis:

Section 3. Discourse functions of spoken backchannel forms

Function	Speaker	Frequency
Continuers (total = 114)	<$1>	17
	<$2>	97
Convergence Tokens (total = 117)	<$1>	56
	<$2>	61
Engaged Response Tokens (total = 11)	<$1>	7
	<$2>	4
Information Receipt Tokens (total = 50)	<$1>	25
	<$2>	25
		292

Section 4. Relationship between head nod type and discourse function

Type	Speaker	CON	CNV	ER	IR	Total
A	<$1>	7	14	1	8	30
	<$2>	12	7	0	4	23
B	<$1>	0	6	2	2	10
	<$2>	39	15	3	2	59
C	<$1>	1	13	1	6	21
	<$2>	6	5	0	5	16
D	<$1>	3	7	1	2	13
	<$2>	8	3	0	2	13
E	<$1>	0	12	1	2	15
	<$2>	13	8	0	1	22
		89	**90**	**9**	**34**	**222**

Section 5. Frequencies of backchanneling head nods types

Type	Speaker	With spoken BC	Without spoken BC	Total
A	<$1>	30	14	44
	<$2>	23	35	58
B	<$1>	9	13	22
	<$2>	59	155	214
C	<$1>	21	11	32
	<$2>	16	11	27
D	<$1>	10	3	13
	<$2>	12	12	24
E	<$1>	8	2	10
	<$2>	19	5	24
	Total	207	261	468

nbc = 6 for <$1> and 4 for <$2>

Appendix 6.6: The analysis of S06FF.

On occasion 1 nod is used across more than one verbal BC turn, details below:

Nod	Speaker	BC no	BC form
D	<$1>	F1d9	Yeah
D	<$1>	F1e1	Mhm
D	<$1>	F1g1	Yes
D	<$1>	F1g2	Yes
D	<$1>	F1g3	Yeah
E	<$1>	F1g6	Yes
E	<$1>	F1g7	Yeah
E	<$1>	F1g8	Yes
E	<$1>	F1g9	Yes
E	<$1>	F1h1	Yes
E	<$1>	F1h7	Right
E	<$1>	F1h8	Yes
E	<$1>	F1h9	Yeah
E	<$1>	F1i6	Yes
E	<$1>	F1i7	Yes
B	<$1>	F1j8	Absolutely yes yes
B	<$1>	F1j9	Absolutely yes
E	<$2>	F2g9	Yeah
E	<$2>	F2h1	Yeah
E	<$2>	F2s3	Yeah
E	<$2>	F2s4	Mhm. That's fine.
E	<$2>	F2s9	Mm
E	<$2>	F2t1	Mhm
D	<$2>	F2t8	Yeah
D	<$2>	F2t9	Mhm

KEY:

	Continuers
	Convergence Tokens
	Engaged Response Tokens
	Information Receipt Tokens
A	Type A nod
B	Type B nod
C	Type C nod
D	Type D nod
E	Type E nod
BC	Spoken Backchannel

Appendix 6.7: The frequencies of specific spoken backchanneling forms and associated functions found in each supervision video.

KEY
- █ = Continuer (Yellow)
- █ = Convergence Token (Cyan)
- █ = Engaged Response Token (Magenta)
- █ = Information Receipt Token (Red)

Note: Each speaker column below is subdivided into four function columns in this order: 1 = Continuer, 2 = Convergence Token, 3 = Engaged Response Token, 4 = Information Receipt Token.

Spoken Backchannel	S01FM.F 1	2	3	4	S01FM.M 1	2	3	4	S02MM.1 1	2	3	4	S02MM.2 1	2	3	4	S03MF.M 1	2	3	4	S03MF.F 1	2	3	4	S04MM.1 1	2	3	4	S04MM.2 1	2	3	4	S05MM.1 1	2	3	4	S05MM.2 1	2	3	4	S06FF.1 1	2	3	4	S06FF.2 1	2	3	4
Absolutely absolutely											1																																					
Absolutely yeah																																			1												
Absolutely yes																																									1						
Absolutely yes yes																																									1						
Ah okay																			1																												
Ah right																																												1			
Ah right. Okay.																				1																											
Ah yes <laughs>																																	1													
Aha																		1																													
Alright								1																																							
Definitely							1								4																						1									1	
Definitely definitely															1																																
Empty yes																															1																
Er																		1																													
Er yes																																	1														
Erm																		1					5			1																						1
Erm yeah																		1																													
Erm yeah yeah okay																						1																									
Exactly yeah er.															1																																
Excellent											1																																				
Excellent. Yes.										1																																					
Good											2																																				
Hm									1				9													1																						
Hm yeah														1																																	
Hmm									1																																						
I agree uh							1																																								
I know																																							1							
I see what you mean																																								1						
Interesting																									1																						
Interesting interesting																									1																						
Interesting isn't it																									1																						
Interesting yeah															1																																
Is it?																								1																							
Is it? Oh																															1																
It does yeah																																								1						
It's excellent											1																																				
Language								1																																							
Mhm																									7				6	1											5			83	1		
Mhm. That's fine.																																													1		
Mm (or Mmm)	103				8				1				243				3				2				104				48				179				89				4				9			
Mm <pause> sure															1																																
Mm mm (or Mmm mmm Mm mmm)	5				1								49												1								3														
Mm mm mm													4																																		
Mm that's right. Yeah.																																	1													
Mm yeah (or Mm. Yeah.)													1																4																		
Mm yeah absolutely																																					1										
Mm yes that's true																																			1												
Mm yes there are quite a few																											1																				
Mm. Interesting isn't it.																											1																				
Mm. Sure.													1																																		
Mm. That's right. Yes. Yeah.														1																																	
No						1	1	1														2			1	2					2		3														
No no					1																													1													
No no but you can						1																																									
No okay																								1																							
No that's right															1																			1													
No that's right no																																		1													
Of sixty																												1																			
Oh does it																				1																											
Oh exactly yeah															1																																
Oh god																				1																											
Oh I know yeah																															1																
Oh I see right																				1																											
Oh I see what you mean yeah																																		1													
Oh is it																											1																				
Oh no not at all no								1																																							
Oh okay																											1																				
Oh okay. Yeah.															1																																
Oh really																			1																												
Oh really? Oh right																				1																											
Oh right								1			2												1																								
Oh right I see																							1																								
Oh right oh okay																			1																												
Oh right okay																							1																								
Oh right yeah																			1					1																							
Oh that should be okay																							1																								
Oh that sort of thing																											1																				
Oh that's interesting								1																																							
Oh wow															1																																
Oh wow. Right.																			1																												
Oh yeah																	1	1									1																				
Ok good																											1																				
Okay (or Ok)	2			3					2				1	5		2	2			6	9			5	4			7				4	1		5				1			1		2			18
Okay brilliant															1																																
Okay I've got you																											1																				
Okay yeah																									2																						
Okay yeah brilliant																							1																								
Okay. Right.															1																																
Okay. Yeah. Mm.																																							1								
Quite interesting																												1																			
Right										2	2	6	1			1	1	5		23	11			3	2				8			8	2			2				2	9		22				6
Right okay (or Right. Okay)				2								3								1																						1						
Right yeah	1																		1	1	1							1																			1
Right yeah I do that yeah																								1																							
Right yeah yeah																				2																											
Right yeah yeah yeah																			1																												
Right. Ah right.																				1																											
Right. Oh right yeah.																				1																											
Simplify																																													1		
Sorry. Yeah.																									1																						
Sure									11	1		1	1	14	1	2											1									2											
Sure yeah yeah																				1																											

Appendix 6.7: The frequencies of specific spoken backchanneling forms and associated functions found in each supervision video.

Spoken Backchannel	S01FM.F				S01FM.M				S02MM.1				S02MM.2				S03MF.M				S03MF.F				S04MM.1				S04MM.2				S05MM.1				S05MM.2				S06FF.1				S06FF.2			
Sure. Sure.													3																																				
That sort of idea. Yeah.																										1																						
That would be ideal yeah																					1																											
That's right. Er.															1																																	
That's okay																																								1								
That's right										1				1				1																3	2												
That's right yeah																																								1								
That's right yes														2																																		
That's right. Yeah yeah.														1																																		
That's true																																	1															
That's true yeah																																	1															
True																																													1			
Uh					1																																											
Uh hm													24																																			
Uh hm that's right yeah														1																																		
Uh huh	12									1								2																														
Uh huh mmm	1																																															
Uhm hm										1																																						
Well that's right exactly																																	1															
Well yeah						1								1																																		
Well yeah absolutely yeah																																			1												
Well yeah yeah																	1																															
Yeah	34	45			2	14			7	9			14	42			40	13			44	30		1	42	41			76	73			15	12			8	33			6	15			5	51		
Yeah <pause> erm																					1																											
Yeah absolutely																																					3											
Yeah definitely																																					1											
Yeah definitely yeah			1																																													
Yeah er (or Yeah err)	1									1											1																											
Yeah erm																				1	1							1																1			
Yeah I know																																													1			
Yeah I know. Yes.																																									1						
Yeah I remember							1																																									
Yeah I think so																					1																											
Yeah it would																					1																											
Yeah mm (or Yeah mmm)	2																								3																							
Yeah no																																	1															
Yeah no go on																											1																					
Yeah no it will do I'm sure			1																																													
Yeah oh god that yeah																		1																														
Yeah okay			2														1								2	3																			1			
Yeah okay mm																										1																						
Yeah right																													1																			
Yeah right good no it looks really good yeah																	1																															
Yeah right yeah																	1																															
Yeah small sample							1																																									
Yeah something like that																		1																														
Yeah sure													3	1	2																																	
Yeah that one yeah																										1																						
Yeah that's a good way to think a link actually yeah							1																																									
Yeah that's right																	1																															
Yeah that's right yeah														1				1																2														
Yeah that's that's interesting														1																																		
Yeah that's true yeah					1																																											
Yeah they do																										1																						
Yeah true																																													1	1		
Yeah twelve I think																						1																										
Yeah uh-huh																		1																														
Yeah we have. Mm.																										1																						
Yeah well yeah																																			1													
Yeah yeah	4	8			1								2				2				1	20			1				1	4			1	9							2				1			
Yeah yeah absolutely																																			1												
Yeah yeah absolutely right														1																																		
Yeah yeah er																	1																															
Yeah yeah of course it is yeah																					1																											
Yeah yeah yeah														1			2																															
Yeah yeah yeah yeah														1																																		
Yeah yeah yeah yeah that's right																					1																											
Yeah yes														1																																		
Yeah you do							1																																									
Yeah, that's important														1																																		
Yeah. Interesting																										1																						
Yeah. Mm okay.																										1																						
Yeah. Right.																		1																														
Yep					3				1				1	1																																		
Yep that's really true							1																																									
Yep yeah	1	6																																														
Yes									2				16								1				25				2				92				1				1	25		1	1			
Yes absoluetly right yes																																											1					
Yes I know																									1																1							
Yes I've got you																									1																							
Yes mhm																																									1							
Yes mm																																	2															
Yes of course yes.														1																																		
Yes that type of thing																									1																							
Yes that's right																																									1							
Yes yeah													2												2	1															1	1						
Yes yeah yeah yeah													1																																			
Yes yes																																	2															
Yes yes yes																																	1															
Yes. Yeah yeah yeah yeah.																									1																						
Yes. Yeah.																													1																			
Yes. Yeah. Mm.													1																																			
Yes. Yes. Yes. Yeah.													1																																			

Appendix 6.8: The frequencies of specific spoken backchanneling forms and associated functions across the five-hour corpus.

Spoken Backchannel	1335	789	131	175	2430
Mm (or Mmm)	793	0	0	0	793
Yeah	293	378	0	1	672
Yes	1	165	0	1	167
Right	1	40	2	73	116
Mhm	101	2	0	0	103
Okay (or Ok)	1	25	0	55	81
Mm. Mm. (or mm mm/ mmm mm	59	0	0	0	59
Yeah yeah	8	49	0	0	57
Sure	1	27	3	3	34
Uh hm	24	0	0	0	24
Uh huh	15	0	0	0	15
No	1	10	1	1	13
Hm	11	0	0	0	11
Yeah okay	0	3	0	6	9
Erm	7	1	0	0	8
That's right	0	0	5	3	8
Definitely	0	0	7	0	7
Right okay (or Right. Okay)	0	0	0	7	7
Yep yeah	1	6	0	0	7
Yes yeah	1	5	0	1	7
Right yeah	0	4	0	2	6
Yeah sure	0	3	1	2	6
Yep	1	5	0	0	6
Mm yeah (or Mm. Yeah.)	0	5	0	0	5
Yeah mm (or Yeah mmm)	2	3	0	0	5
Mm mm mm	4	0	0	0	4
Oh right	0	0	2	2	4
Yeah erm	1	3	0	0	4
Yeah that's right yeah	0	0	4	0	4
Oh yeah	0	1	2	0	3
Sure. Sure.	0	3	0	0	3
Yeah absolutely	0	0	3	0	3
Yeah er (or Yeah err)	1	2	0	0	3
Yeah yeah yeah	0	3	0	0	3
Aha	1	0	1	0	2
Good	0	0	2	0	2
No no	1	1	0	0	2
No that's right	0	0	2	0	2
Okay yeah	0	2	0	0	2
Right yeah yeah	0	0	0	2	2
That's right yes	0	0	2	0	2
Well yeah	0	1	1	0	2
Yeah true	0	1	1	0	2
Yes I know	0	1	1	0	2
Yes mm	0	2	0	0	2
Yes yes	0	2	0	0	2
Absolutely absolutely	0	0	1	0	1
Absolutely yeah	0	0	1	0	1
Absolutely yes	0	0	1	0	1
Absolutely yes yes	0	0	1	0	1
Ah okay	0	0	1	0	1
Ah right	0	0	0	1	1
Ah right. Okay.	0	0	0	1	1
Ah yes <laughs>	0	0	1	0	1
Alright	0	0	0	1	1
Definitely definitely	0	0	1	0	1
Empty yes	0	0	1	0	1
Er	1	0	0	0	1
Er yes	0	1	0	0	1
Erm yeah	1	0	0	0	1
Erm yeah yeah okay	0	0	1	0	1
Exactly yeah er.	0	0	1	0	1
Excellent	0	0	1	0	1
Excellent. Yes.	0	1	0	0	1
Hm yeah	0	1	0	0	1
Hmm	1	0	0	0	1
I agree uh	0	1	0	0	1
I know	0	0	1	0	1
I see what you mean	0	1	0	0	1
Interesting	0	0	1	0	1
Interesting interesting	0	0	1	0	1
Interesting isn't it	0	0	1	0	1
Interesting yeah	0	0	1	0	1
Is it?	0	0	1	0	1
Is it? Oh	0	0	1	0	1
It does yeah	0	0	1	0	1
It's excellent	0	0	1	0	1
Language	0	0	1	0	1
Mhm. That's fine.	0	1	0	0	1
Mm <pause> sure	0	1	0	0	1
Mm that's right. Yeah.	0	0	1	0	1
Mm yeah absolutely	0	0	1	0	1
Mm yes that's true	0	0	1	0	1
Mm yes there are quite a few	0	0	1	0	1
Mm. Interesting isn't it.	0	0	1	0	1
Mm. Sure.	0	1	0	0	1
Mm. That's right. Yes. Yeah.	0	0	1	0	1
No no but you can	0	0	1	0	1
No okay	0	0	0	1	1
No that's right no	0	0	1	0	1
Of sixty	0	0	1	0	1
Oh does it	0	0	1	0	1
Oh exactly yeah	0	0	1	0	1
Oh god	0	0	1	0	1
Oh I know yeah	0	0	1	0	1
Oh I see right	0	0	1	0	1
Oh I see what you mean yeah	0	0	1	0	1
Oh is it	0	0	1	0	1

Spoken Backchannel	1335	789	131	175	2430
Oh no not at all no	0	0	0	1	1
Oh okay	0	0	1	0	1
Oh okay. Yeah.	0	0	1	0	1
Oh really	0	0	1	0	1
Oh really? Oh right	0	0	1	0	1
Oh right I see	0	0	1	0	1
Oh right oh okay	0	0	0	1	1
Oh right okay	0	0	1	0	1
Oh right yeah	0	0	1	0	1
Oh that should be okay	0	0	1	0	1
Oh that sort of thing	0	0	1	0	1
Oh that's interesting	0	0	1	0	1
Oh wow	0	0	1	0	1
Oh wow. Right.	0	0	1	0	1
Ok good	0	0	1	0	1
Okay brilliant	0	0	0	1	1
Okay I've got you	0	0	1	0	1
Okay yeah brilliant	0	0	0	1	1
Okay. Right.	0	0	0	1	1
Okay. Yeah. Mm.	0	0	0	1	1
Quite interesting	0	0	1	0	1
Right yeah I do that yeah	0	0	1	0	1
Right yeah yeah yeah	0	1	0	0	1
Right. Ah right.	0	0	0	1	1
Right. Oh right yeah.	0	0	0	1	1
Simplify	0	0	1	0	1
Sorry. Yeah.	0	1	0	0	1
Sure yeah yeah	0	0	1	0	1
That sort of idea. Yeah.	0	0	1	0	1
That would be ideal yeah	0	0	1	0	1
That's right. Er.	0	0	0	1	1
That's okay	0	0	1	0	1
That's right yeah	0	0	1	0	1
That's right. Yeah yeah.	0	0	1	0	1
That's true	0	0	1	0	1
That's true yeah	0	0	1	0	1
True	0	0	1	0	1
Uh	1	0	0	0	1
Uh hm that's right yeah	0	0	1	0	1
Uh huh mmm	1	0	0	0	1
Uhm hm	1	0	0	0	1
Well that's right exactly	0	0	1	0	1
Well yeah absolutely yeah	0	1	0	0	1
Well yeah yeah	0	1	0	0	1
Yeah <pause> erm	0	1	0	0	1
Yeah definitely	0	0	1	0	1
Yeah definitely yeah	0	0	1	0	1
Yeah I know	0	1	0	0	1
Yeah I know. Yes.	0	1	0	0	1
Yeah I remember	0	0	1	0	1
Yeah I think so	0	1	0	0	1
Yeah it would	0	1	0	0	1
Yeah no	0	1	0	0	1
Yeah no go on	0	0	1	0	1
Yeah no it will do I'm sure	0	0	1	0	1
Yeah oh god that yeah	0	0	1	0	1
Yeah okay mm	0	0	1	0	1
Yeah right	0	0	0	1	1
Yeah right good no it looks really	0	0	0	1	1
Yeah right yeah	0	0	0	1	1
Yeah small sample	0	0	1	0	1
Yeah something like that	0	0	1	0	1
Yeah that one yeah	0	0	1	0	1
Yeah that's a good way to think	0	0	1	0	1
Yeah that's right	0	1	0	0	1
Yeah that's that's interesting	0	0	1	0	1
Yeah that's true yeah	0	1	0	0	1
Yeah they do	0	0	1	0	1
Yeah twelve I think	0	0	1	0	1
Yeah uh-huh	0	1	0	0	1
Yeah we have. Mm.	0	0	1	0	1
Yeah well yeah	0	1	0	0	1
Yeah yeah absolutely	0	0	1	0	1
Yeah yeah absolutely right	0	0	1	0	1
Yeah yeah er	0	1	0	0	1
Yeah yeah of course it is yeah	0	0	1	0	1
Yeah yeah yeah yeah	0	1	0	0	1
Yeah yeah yeah yeah that's right	0	0	1	0	1
Yeah yes	0	1	0	0	1
Yeah you do	0	1	0	0	1
Yeah, that's important	0	0	1	0	1
Yeah. Interesting	0	0	1	0	1
Yeah. Mm okay.	0	0	1	0	1
Yeah. Right.	0	1	0	0	1
Yep that's really true	0	0	1	0	1
Yes absoluetly right yes	0	0	1	0	1
Yes I've got you	0	0	1	0	1
Yes mhm	0	1	0	0	1
Yes of course yes.	0	1	0	0	1
Yes that type of thing	0	0	1	0	1
Yes that's right	0	1	0	0	1
Yes yeah yeah yeah	0	1	0	0	1
Yes yes yes	0	1	0	0	1
Yes. Yeah yeah yeah yeah.	0	1	0	0	1
Yes. Yeah.	0	1	0	0	1
Yes. Yeah. Mm.	0	1	0	0	1
Yes. Yes. Yes. Yeah.	0	1	0	0	1

KEY

- 1335 = Continuer
- 789 = Convergence Token
- 131 = Engaged Response Token
- 175 = Information Recipt Token
- 2430 = Total Frequency (across all functions)

Appendix 6.9: The frequencies of specific spoken backchanneling forms / functions and concurrent backchanneling head nod types in each supervision.

Spoken Backchannel	S01FM.F					S01FM.M					S02MM.1					S02MM.2					S03MF.M					S03MF.F				
	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E
Absolutely absolutely												1																		
Ah yes <laughs>																					1									
Definitely							1										1			1										
Erm yeah yeah okay																						1								
Excellent											1																			
Excellent. Yes.											1																			
Good																				1										
Hm														1		2	5													
Hm yeah																				1										
It's excellent											1																			
Mm (or Mmm)	19	21	7	2	1	3					1					47	83	13	12	48	1					1				
Mm <pause> sure																		1												
Mm mm (or Mmm mmm or Mm. Mm.)		2			2											6	19	4	2	13										
Mm mm mm																1	1			1										
Mm. Sure.																	1													
Mm. That's right. Yes. Yeah.																		1												
No	1					1																							1	
No okay																										1				
No that's right																1														
Oh does it																		1												
Oh okay. Yeah.																1														
Oh right			1								1																		1	
Oh right I see																													1	
Oh right okay																												1		
Oh right yeah																								1						
Oh that should be okay																										1				
Oh yeah																								1						
Okay (or Ok)	1		1			1					4	1			1					1	3		3			7		1	3	
Okay brilliant																							1							
Okay yeah brilliant																					1									
Okay. Right.													1																	
Right												1	6					1			4	4	6		1	6		4		
Right okay (or Right. Okay)			1								1	1		1							1									
Right yeah			1																		1		1			1				
Right yeah I do that yeah																												1		
Right. Ah right.																							1							
Right. Oh right yeah.																					1									
Sure											7	1	1		1	4	4	3	1	4										
Sure yeah yeah																	1				1									
Sure. Sure.																1				1										
That would be ideal yeah																										1				
That's right. Er.																		1												
That's right														1																
That's right. Yeah yeah.																		1												
That's right yes																				1										
Uh hm																8	3	2		4										
Uh hm that's right yeah																		1												
Uh huh	3	2	3								1												1	1						
Uhm hm											1																			
Well yeah yeah																							1							
Yeah	18	15	16	4	2	9	2	1			5		1		3	9	14	7	1	10	8	15	5		4	37	5	22	1	2
Yeah <pause> erm																										1				
Yeah er (or Yeah err)				1																								1		
Yeah erm																										1				
Yeah I think so																												1		
Yeah it would																												1		
Yeah no it will do I'm sure		1																												
Yeah okay	1																				1									
Yeah right good no it looks really good yeah																						1								
Yeah sure																1				4										
Yeah that's a good way to think a link actually		1																												
Yeah that's right																							1							
Yeah that's right yeah																								1						
Yeah uh-huh																													1	
Yeah yeah	3		5	2	2		1				1	1						1		1	5	7	2	2	1				1	
Yeah yeah absolutely right																				1										
Yeah yeah er																					1									
Yeah yeah of course it is yeah																					1									
Yeah yeah yeah																		1			1	1								
Yeah yeah yeah yeah																							1							
Yeah yeah yeah yeah that's right																													1	
Yeah yes																				1										
Yeah you do				1																										
Yeah. Right.																							1							
Yep	1	1	1						1			1								1										
Yep that's really true	1																													
Yep yeah			6																											
Yes												1			1	4	4	5		1								1		
Yes of course yes.																	1													
Yes yeah																1		1												
Yes yeah yeah yeah																				1										
Yes. Yes. Yes. Yeah.																				1										

Appendix 6.9: The frequencies of specific spoken backchanneling forms / functions and concurrent backchanneling head nod types in each supervision.

Spoken Backchannel	S04MM.1					S04MM.2					S05MM.1					S05MM.2					S06FF.1					S06FF.2				
	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E
Absolutely yeah											1																			
Absolutely yes																						1								
Absolutely yes yes																						1								
Empty yes							1																							
Er yes													1																	
Hm	1																													
I know																									1					
I see what you mean																							1							
Interesting interesting	1																													
It does yeah																					1									
Mhm		1	3			5	1														2			1		8	35	4	8	12
Mhm. That's fine.																														1
Mm (or Mmm)	29	11	11	8	5	19	16				85	50	4		1	56	2	10			3					2	3	1	1	1
Mm mm (or Mmm mmm)											3																			
Mm that's right. Yeah.												1																		
Mm yeah (or Mm. Yeah.)						3		1																						
Mm yeah absolutely																1														
Mm yes that's true															1															
Mm yes there are quite a few		1																												
Mm. Interesting isn't it.			1																											
No						2					1	1																		
No no											1																			
No that's right no											1																			
Of sixty													1																	
Oh I see what you mean yeah											1																			
Oh okay	1																													
Oh yeah			1																											
Ok good	1																													
Okay (or Ok)	2		2	3							2	1	2						1							4	2	3	1	1
Okay I've got you			1																											
Okay yeah	1		1																											
Okay. Yeah. Mm.			1																											
Right		1				5		4			1					1					12	1	9	2	2			3	1	
Right okay (or Right. Okay)																		1												
Simplify																											1			
Sorry. Yeah.			1																											
Sure	1										1		1																	
That sort of idea. Yeah.	1																													
That's right												1	1	1	1															
That's right yeah																							1							
That's true															1															
That's true yeah											1																			
True																											1			
Well that's right exactly												1																		
Well yeah absolutely yeah																		1												
Yeah	14	12	6	14	10	44	33	11	3	2	7	5	5			16	4	10			5	1	6	2	3	9	15	5	2	7
Yeah absolutely																2		1												
Yeah definitely																		1												
Yeah I know. Yes.																							1							
Yeah no												1																		
Yeah okay	1			1	1																						1			
Yeah okay mm		1																												
Yeah right			1																											
Yeah that's right											1																			
Yeah they do		1																												
Yeah true																											1			
Yeah we have. Mm.				1																										
Yeah well yeah																			1											
Yeah yeah	1	2	1			3	3		1							2						1								
Yeah yeah absolutely							1																							
Yeah yes												1																		
Yeah. Interesting	1																													
Yeah. Mm okay.			1																											
Yes	7	3	1	4	1	2					36	14	12	2	4						6	2	4	6	8					
Yes absolutely right yes																								1						
Yes I know																						1								
Yes I've got you		1																			1									
Yes mm											1	1																		
Yes that type of thing		1																												
Yes that's right																						1								
Yes yeah		1		1																				1	1					
Yes yes											1	1																		
Yes. Yeah yeah yeah yeah.			1																											

Appendix 6.10: The frequencies of specific spoken backchanneling forms / functions and concurrent backchanneling head nod types. across the five-hour corpus.

Spoken Backchannel	A	B	C	D	E
Absolutely absolutely	0	1	0	0	0
Absolutely yeah	1	0	0	0	0
Absolutely yes	0	1	0	0	0
Absolutely yes yes	0	1	0	0	0
Aha	1	0	0	0	0
Definitely	0	2	0	0	1
Empty yes	0	1	0	0	0
Er yes	0	0	1	0	0
Erm yeah yeah okay	0	1	0	0	0
Excellent	1	0	0	0	0
Excellent. Yes.	1	0	0	0	0
Good	0	0	0	0	1
Hm	3	5	0	1	0
Hm yeah	0	0	0	0	1
I know	0	0	0	0	1
I see what you mean	0	0	1	0	0
Interesting interesting	1	0	0	0	0
It does yeah	1	0	0	0	0
It's excellent	1	0	0	0	0
Mhm	15	37	7	9	12
Mhm. That's fine.	0	0	0	0	1
Mm (or mmm)	266	186	46	23	56
Mm <pause> sure	0	0	1	0	0
Mm mm (or mmm mm or mmm mmm)	9	21	4	2	15
Mm mm mm	1	1	0	0	1
Mm that's right. Yeah.	0	1	0	0	0
Mm yeah	3	0	1	0	0
Mm yeah absolutely	1	0	0	0	0
Mm yes that's true	0	0	0	0	1
Mm yes there are quite a few	0	1	0	0	0
Mm. Interesting isn't it.	0	0	1	0	0
Mm. Sure.	0	1	0	0	0
Mm. That's right. Yes. Yeah.	0	0	1	0	0
No	5	1	1	0	0
No no	1	0	0	0	0
No okay	1	0	0	0	0
No that's right	1	1	0	0	0
No that's right no	1	0	0	0	0
Oh exactly yeah	0	0	1	0	0
Oh I see what you mean yeah	1	0	0	0	0
Oh okay	1	0	0	0	0
Oh okay. Yeah.	1	0	0	0	0
Oh right	1	0	2	0	0
Oh right I see	0	0	1	0	0
Oh right okay	1	0	0	0	0
Oh right yeah	0	0	0	0	1
Oh that should be okay	0	1	0	0	0
Oh yeah	0	0	1	0	1
Ok good	1	0	0	0	0
Okay	24	5	15	4	3
Okay brilliant	0	0	1	0	0
Okay I've got you	0	0	0	1	0
Okay yeah	1	0	0	1	0
Okay yeah brilliant	1	0	0	0	0
Okay. Right.	0	0	1	0	0
Okay. Yeah. Mm.	0	0	0	1	0
Right	29	8	32	3	3
Right okay	2	1	2	1	0
Right yeah	2	0	2	0	0
Right yeah I do that yeah	0	1	0	0	0
Right yeah yeah yeah	0	1	0	0	0
Right. Oh right yeah.	1	0	0	0	0
Simplify	0	1	0	0	0
Sorry. Yeah.	0	0	1	0	0
Sure	13	5	5	1	5
Sure yeah yeah	1	0	0	0	0
Sure. Sure.	1	1	0	0	1
That sort of idea. Yeah.	1	0	0	0	0
That would be ideal yeah	1	0	0	0	0
That's right. Er.	0	0	1	0	0
That's right	0	1	1	2	1
That's right yeah	0	0	1	0	0
That's right yes	0	0	0	0	1
That's right. Yeah yeah.	0	0	1	0	0
That's true	0	0	0	0	1
That's true yeah	1	0	0	0	0
True	0	1	0	0	1
Uh hm	8	3	2	0	4
Uh hm that's right yeah	0	0	1	0	0
Uh-huh	4	2	4	1	0
Well that's right exactly	0	1	0	0	0
Well yeah absolutely yeah	0	0	1	0	0
Well yeah yeah	0	1	0	0	0
Yeah	182	120	96	31	39
Yeah <pause> erm	1	0	0	0	0
Yeah absolutely	2	0	1	0	0
Yeah definitely	0	0	1	0	0
Yeah er	0	0	1	0	0
Yeah erm	1	0	0	0	0
Yeah err	0	0	0	0	1
Yeah I know. Yes.	0	1	0	0	0
Yeah I think so	0	0	1	0	0
Yeah it would	0	0	1	0	0
Yeah no	0	0	1	0	0
Yeah no it will do I'm sure	0	0	1	0	0
Yeah okay	3	1	0	1	1
Yeah okay mm	0	1	0	0	0
Yeah right	0	1	0	1	0
Yeah right good no it looks really good yeah	0	1	0	0	0

Spoken Backchannel	A	B	C	D	E
Yeah sure	1	0	0	0	4
Yeah Sure	1	0	0	0	4
Yeah that's a good way to think a link actually	0	0	1	0	0
Yeah that's right	1	0	1	0	0
Yeah that's right yeah	0	0	0	1	0
Yeah they do	0	1	0	0	0
Yeah true	0	1	0	30	0
Yeah uh-huh	0	0	1	0	0
Yeah we have. Mm.	0	0	0	0	1
Yeah well yeah	0	0	0	0	1
Yeah yeah	15	16	9	5	4
Yeah yeah absolutely	0	1	0	0	0
Yeah yeah absolutely right	0	0	0	0	1
Yeah yeah er	1	0	0	0	0
Yeah yeah of course it is yeah	1	0	0	0	0
Yeah yeah yeah	1	2	0	0	0
Yeah yeah yeah yeah	0	1	0	0	0
Yeah yeah yeah yeah that's right	0	0	0	0	1
Yeah yes	0	0	0	0	1
Yeah you do	0	0	0	0	1
Yeah. Interesting	1	0	0	0	0
Yeah. Mm okay.	0	0	0	1	0
Yep	1	1	3	0	1
Yep that's really true	1	0	0	0	0
Yep yeah	0	0	6	0	0
Yes	55	23	24	12	15
Yes absolutely right yes	0	0	0	1	0
Yes I know	0	1	0	0	0
Yes I've got you	0	1	0	0	0
Yes mhm	1	0	0	0	0
Yes mm	1	1	0	0	0
Yes of course yes.	0	1	0	0	0
Yes that type of thing	0	1	0	0	0
Yes that's right	0	1	0	0	0
Yes yeah	1	1	1	2	1
Yes yeah yeah yeah	0	0	1	0	0
Yes yes	1	1	0	0	0
Yes yes yes	0	1	0	0	0
Yes. Yeah yeah yeah yeah.	0	0	0	1	0
Yes. Yes. Yes. Yeah.	0	0	0	0	1

Overall = 1737 spoken backchannels with concurrent nods

Individual Totals:

A	B	C	D	E
682	476	288	105	186

KEY

A	=	Nod Type A
B	=	Nod Type B
C	=	Nod Type C
D	=	Nod Type D
E	=	Nod Type E

357

Appendix 6.11:
Exploring the use of backchanneling head nods and concurrent spoken backchannel forms across turns
(charting individual speakers/ videos).

Section 1: S01FM

Section 2: S02MM

Section 3: S03MF

Section 4: S04MM

358

KEY:

Backchanneling nod:

bf = Starts before concurrent BC ends with BC

af = Starts with concurrent BC, proceeds after it

a&b = Starts before BC is uttered and proceeds after it

BC = Spoken backchannel
A = Type A backchanneling nod
D = Type B backchanneling nod
C = Type C backchanneling nod
D = Type D backchanneling nod
E = Type E backchanneling nod

= Continuer
= Convergence Token
= Engaged Response Token
= Information Receipt Token

Appendix 6.11:
Exploring the use of backchanneling head nods and concurrent spoken backchannel forms across turns
(charting individual speakers/ videos).

Section 5: S05MM

5a. S05MM.1

BC Form and Function →

Location and type of nod (in relation to BC) ↓

		Mm	Mm that's right. Yeah.	Mm yes that's true	No	Okay	Sure	That's right	That's right	That's true	Well that's right exactly	Yeah	Yeah	Yes	Yes yes	Yes yes yes	Σ
bf sum: 35	A	1															1
	B	16	1		1	1						2		2	5		28
	C																0
	D													1			1
	E	1												4			5
af sum: 42	A	2															2
	B	24				1		1				7	1	1			35
	C													1			1
	D				1			1						1			1
	E													1			3
a&b sum: 12	A																0
	B	8									1	1					10
	C			1													1
	D																0
	E							1									1
		52	1	1	1	1	1	1	1		2	3	21	1	1		**89**

5b. S05MM.2

BC Form and Function →

Location and type of nod (in relation to BC) ↓

		mm	yeah	yeah	Σ
bf sum: 5	A	2	1		3
	B			2	2
	C				0
	D				0
	E				0
af sum: 3	A	1	1		1
	B	1	1		2
	C				0
	D				0
	E				0
a&b sum: 1	A				0
	B	1			1
	C				0
	D				0
	E				0
		4	3	2	**9**

Section 2: S06FF

6a. S06FF.1

BC Form and Function →

Location and type of nod (in relation to BC) ↓

		Absolutely yes	Absolutely yes yes	It does yeah	Mhm	Right	That's right yeah	Yeah	Yeah I know. Yes.	Yes	Yes	Yes	Yes I know	Yes that's right	Yes yeah	Yes yeah	Σ
bf sum: 14	A	1					1					1	1				4
	B			1	1												2
	C			1	1												5
	D																2
	E									1		1					2
af sum: 12	A	1			1												2
	B				1		1	1									3
	C				1					1							2
	D				1		1			1		1					4
	E																1
a&b sum: 15	A																0
	B							1									1
	C																0
	D			1					2								3
	E			2	2			7									11
		1	1	1	6	1	9	1	1	1	14	1	1	1	1		**41**

6b. S06FF.2

BC Form and Function →

Location and type of nod (in relation to BC) ↓

		Mhm	Mhm	Mhm. That's fine.	Mm	Okay	Right	Simplify	True	Yeah	Yeah	Yeah true	Σ
bf sum: 65	A												0
	B	27		2	2			1	13	1	1		47
	C												0
	D	2	1		1	1			1				6
	E	8		1					3				12
af sum: 12	A												0
	B	4				1		1					6
	C	1							1				1
	D								1				2
	E	2		1									3
a&b sum: 16	A												0
	B	3	1										4
	C												0
	D	4		1				1					5
	E	2						4					7
		54	1	1	5	4	1	1	23	1		1	**93**

Appendix 6.12

Table 1 (Mm/Mmm family)

BC Form / Function	Total	bf (Total = 228) A B C D E	af (Total = 128) A B C D E	b&a (Total = 110) A B C D E
Absolutely yes	1	1		
Absolutely yes yes	1		1	
Definitely	3	1	1	1
Empty yes	1		1	
Excellent	1			
Good	1	1		
Hm	6	2 2	1	1
It does yeah	1		1	
Mhm	59	28 1 8 3 1	5 1	2 3 1 4 2
Mhm	1			
Mhm. That's fine.	1			
Mm / Mmm	280	13 87 21 8 3	4 56 7 2	11 37 6 1 25
Mm mm / Mmm mmm	35	1 3	5	6 1 7
Mm that's right. Yeah.	1	1		
Mm yes that's true	1		1	
Mm. Mm. Mm.	1			1
Mm. Sure.	1		1	
No	1			1
Oh right yeah	1	1		
Okay	1	1		
Okay	2	1		1
Okay	9	4	1 2 1	1
Okay I've got you	1		1	
Okay yeah	1	1		
Right	3	1 1 1		1
Right	15	1 5 1	2 1	2 1 3
Right Okay	2	1	1 1	
Simplify	1		1	
Sure	11	4	1	3 1
Sure	1		1	
Sure. Sure.	1	1		
That's right. Er.	1	1		
That's right	2			1 1
That's right	1		1	
That's right yeah	1	1		
that's right yes	1			1
That's right. Yeah yeah.	1	1		
That's true	1		1	
True	1	1		
Uh hm	8		1 1	3
Uh-huh	2		1	1
Well that's right exactly	1	1		
well yeah yeah	1	1		
TOTAL	**466**	**16 141 13 18 40**	**8 78 3 15 24**	**0 51 2 12 45**

Table 2 (Yeah/Yes family)

BC Form / Function	Total	bf (Total = 154) A B C D E	af (Total = 79) A B C D E	b&a (Total = 68) A B C D E
Yeah	91	4 5 30 33 3	2 8 1 4 5	14 1 2 6
Yeah	101	7 6 9 2 1	7 4 4 8	3 3 11
yeah err	1	1		
Yeah I know. Yes.	1	1		
Yeah right	1	1		
Yeah. Right.	1	1		
Yeah sure	2		1	1
Yeah sure	1		1	1
Yeah. Sure.	1		1	
yeah they do	1	1		
Yeah true	1	1		
Yeah we have. Mm.	1		1	
yeah yeah	17	8	3 2	1 1 1
Yeah yeah	4	2 1	2	
Yeah yeah absolutely right	1			1
Yeah yeah yeah	1	1		
yeah yeah yeah yeah	1	1		
yeah yeah yeah yeah that's right	1		1	
Yeah yes	1		1	
yeah you do	1		1	
Yeah. Mm okay.	1		1	
Yep	2		1 1	
Yep	1		1	
yep yeah	1		1 1	
Yes	51	5 9 2 4 2	9 3 4 1 3	2 9
Yes	1		1	
Yes	1		1	
Yes	0			
Yes I know	1	1		
yes I've got you	1	1		
Yes okay	2		1	
Yes that's right	1	1		
Yes yeah	1	1		
Yes yeah	3		1	
Yes yeah yeah yeah	1	1		
Yes yes	1	1 1		
Yes yes yes	1	1		
Yes. Yes. Yes. Yeah.	1	1		
	301	**9 89 6 24 26**	**2 39 5 14 19**	**0 26 3 10 29**
GRAND TOTAL	**767**	**25 230 19 42 66**	**10 117 8 29 43**	**0 77 5 22 74**

KEY:

Backchanneling nod:
- **bf** = Starts before BC, ends with BC
- **af** = Starts with concurrent BC, proceeds after it
- **b&a** = Starts before BC is uttered and proceeds after it

- **BC** = Spoken backchannel
- **A** = Type A backchanneling nod
- **B** = Type B backchanneling nod
- **C** = Type C backchanneling nod
- **D** = Type D backchanneling nod
- **E** = Type E backchanneling nod

- Red = Convergence Token
- Magenta = Continuer
- Cyan = Engaged Response Token
- Yellow = Information Receipt Token

Appendix 6.12: Exploring the use of backchanneling head nods and concurrent spoken backchannel forms across turns (combining results from all videos).

Appendix 6.13: Scatter plots representing the use of spoken and non-verbal backchannels across each video (and speaker) in the five-hour corpus.

| - denotes the approximate time at which a specific backchannel was used across each conversation

Plot 1

A scatter plot of all spoken and non-verbal backchannels used by each speaker in the sub-corpus.

Start of conversation   TIME   End of conversation
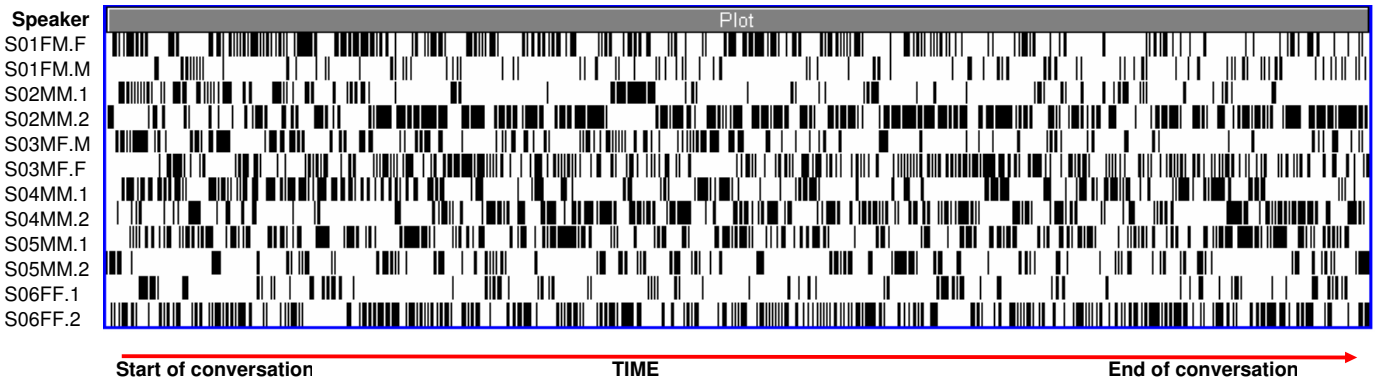
Plot 2

A scatter plot of all spoken backchannels used in the sub-corpus (with and without concurrent backchanneling nods).

Start of conversation   TIME   End of conversation
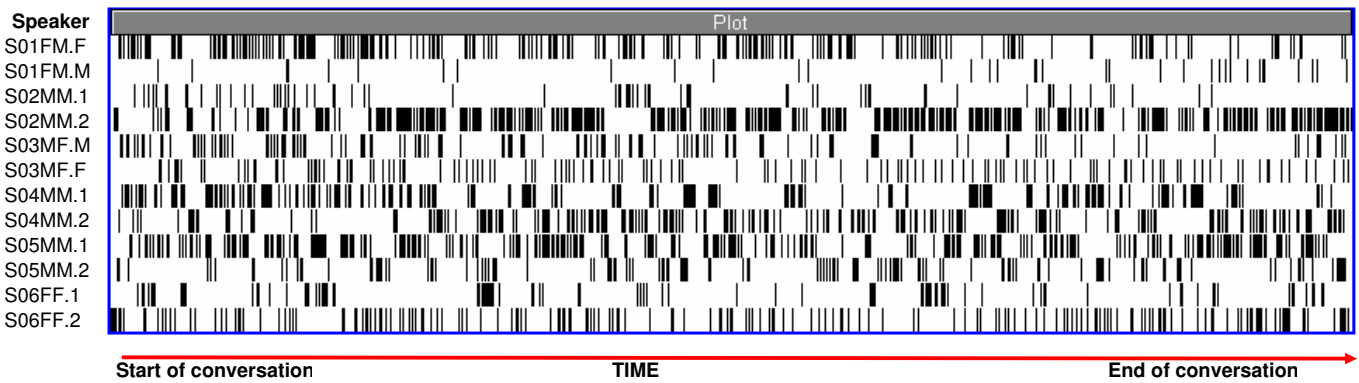
Plot 3

A scatter plot of all backchanneling nods used in the sub-corpus (with and without concurrent spoken backchannels).

Start of conversation   TIME   End of conversation
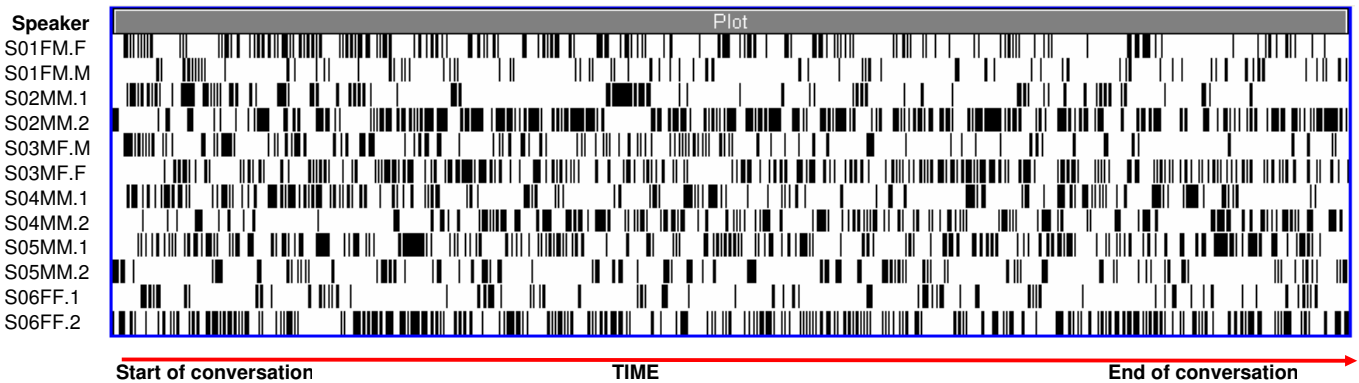
Plot 4

A scatter plot of backchanneling nods used without conurrent spoken backchannels.
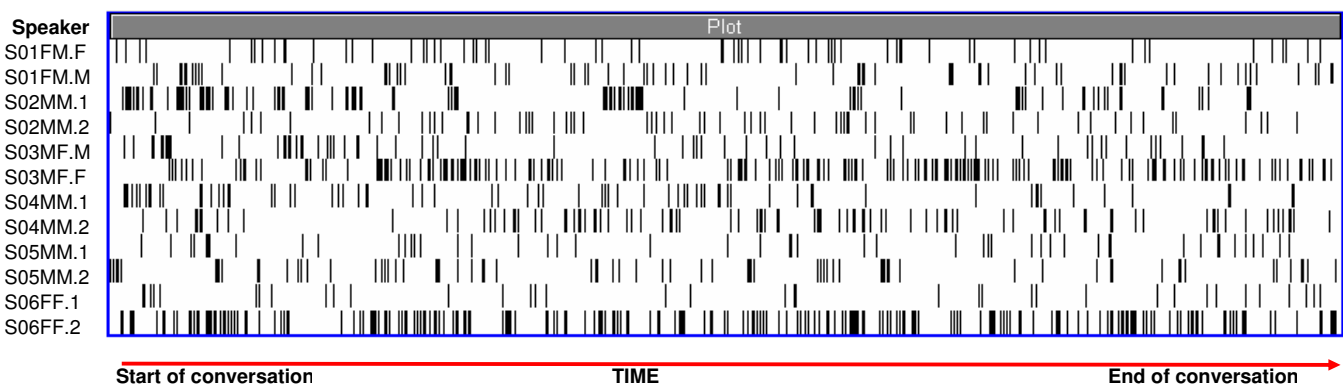
Start of conversation   TIME   End of conversation

Scatter plots representing the use of spoken and non-verbal backchannels across each video (and speaker) in the five-hour corpus.

Plot 5

A scatter plot of backchanneling type A nods, used with conurrent spoken backchannels.
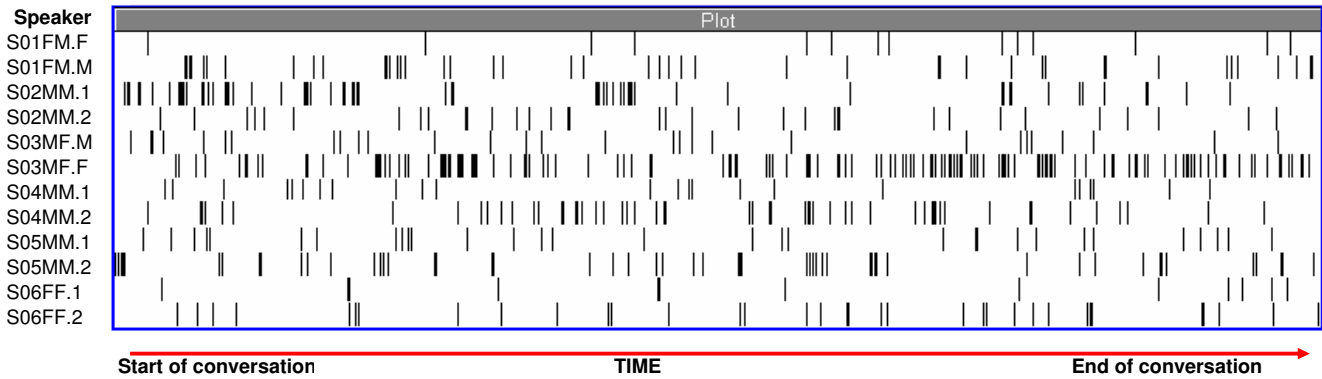
Plot 6

A scatter plot of backchanneling type B nods, used with conurrent spoken backchannels.
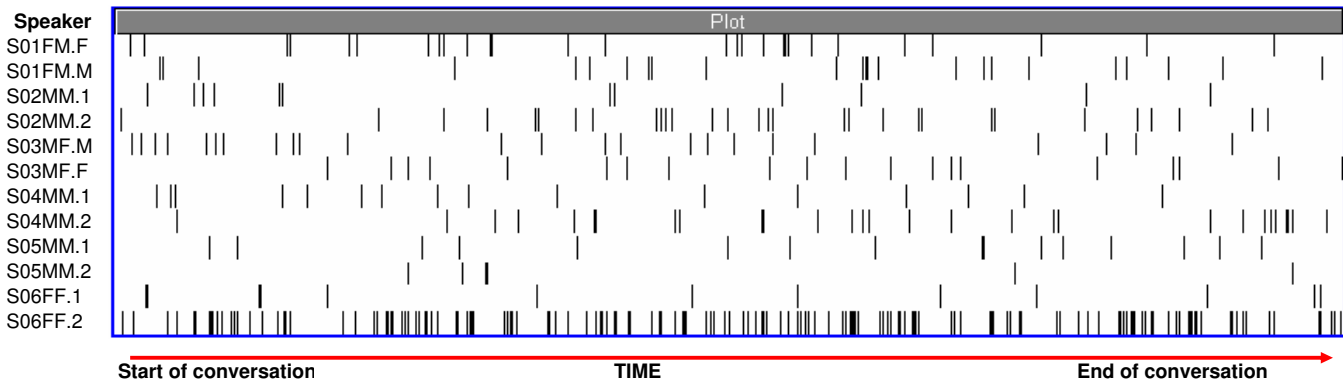
Plot 7

A scatter plot of backchanneling type C nods, used with conurrent spoken backchannels.
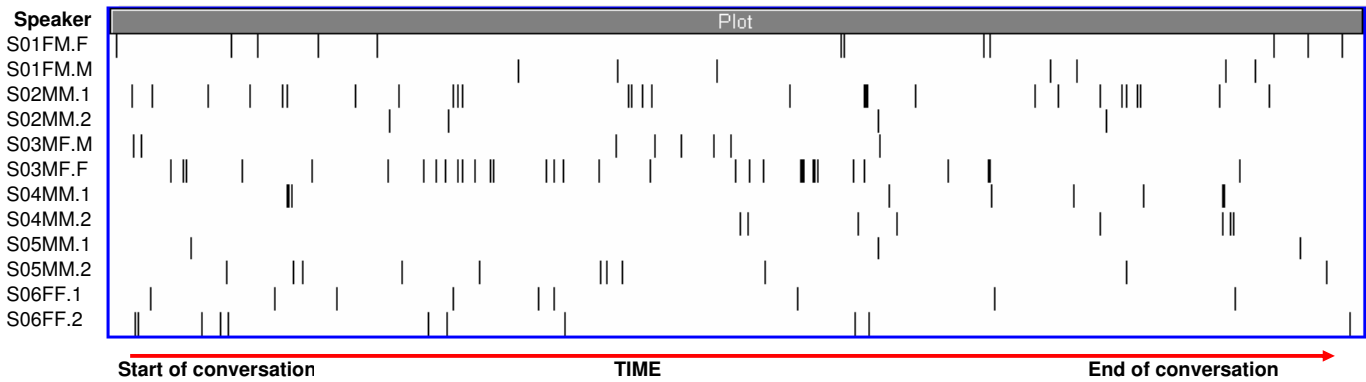
Plot 8

A scatter plot of backchanneling type D nods, used with conurrent spoken backchannels.
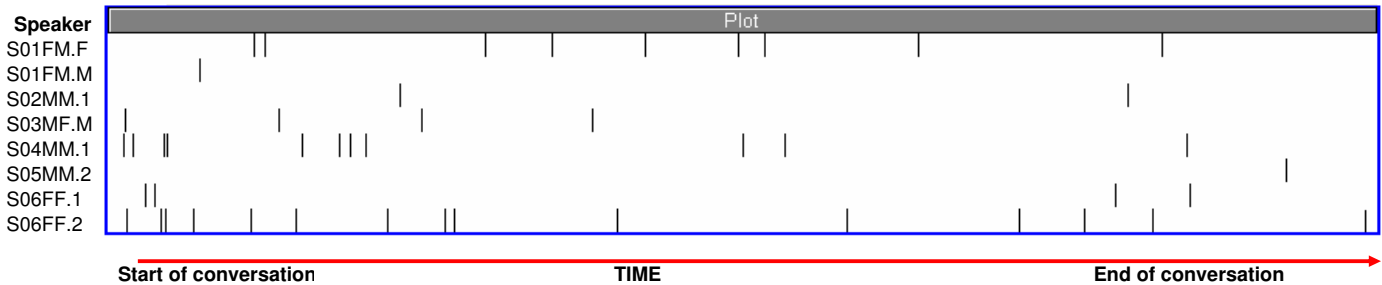
Appendix 6.13: Scatter plots representing the use of spoken and non-verbal backchannels across each video (and speaker) in the five-hour corpus.

Plot 9
A scatter plot of backchanneling type E nods, used with conurrent spoken backchannels.
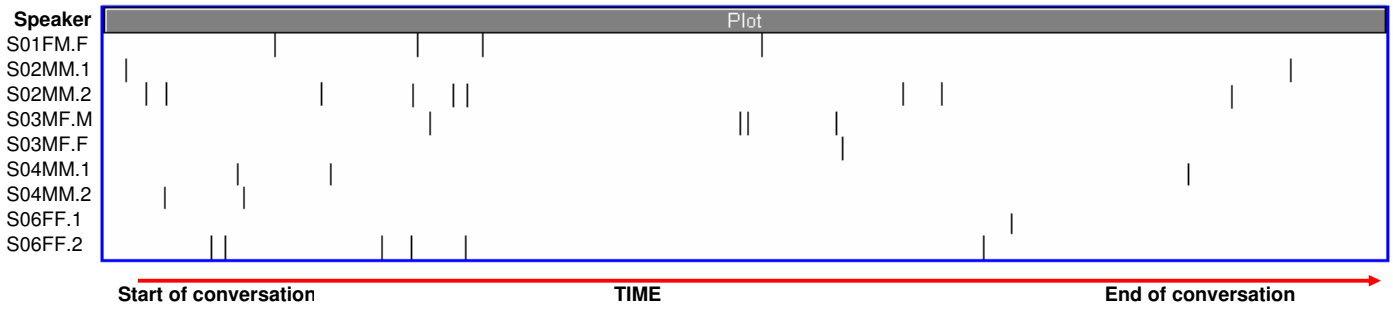
Plot 10
A scatter plot of spoken backchannels, used without conurrent backchanneling nods.
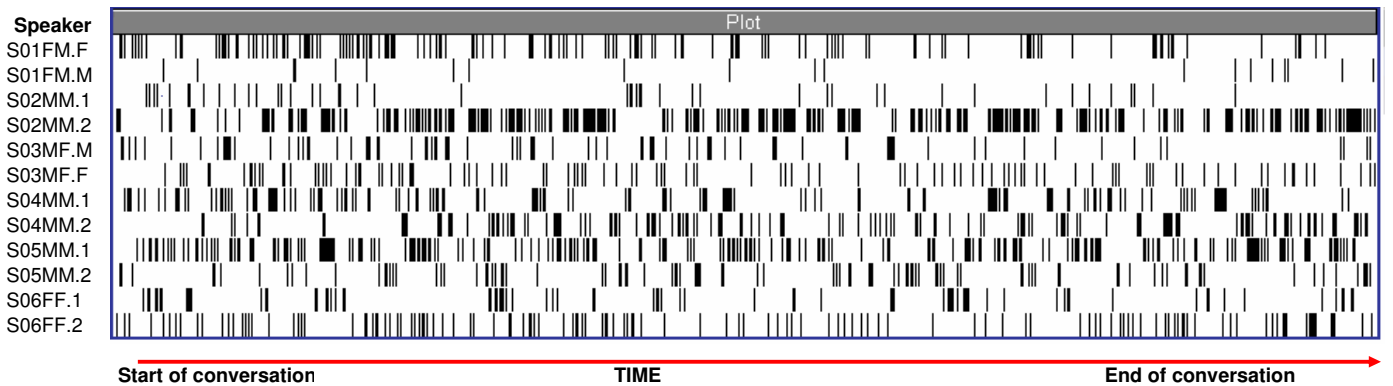
Plot 11
A scatter plot of spoken backchannels used with conurrent backchanneling nods.
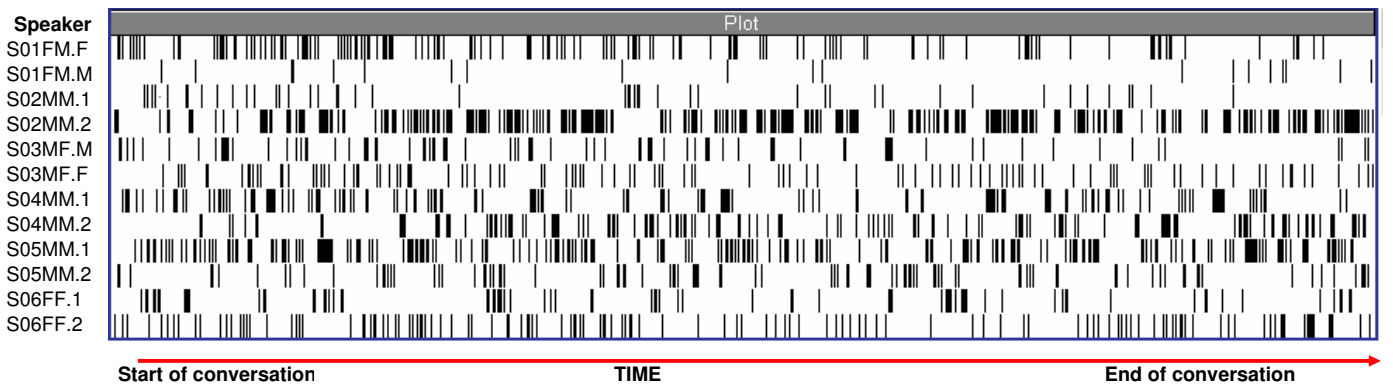
Plot 12
A scatter plot of spoken backchannels used with conurrent type A backchanneling nods.
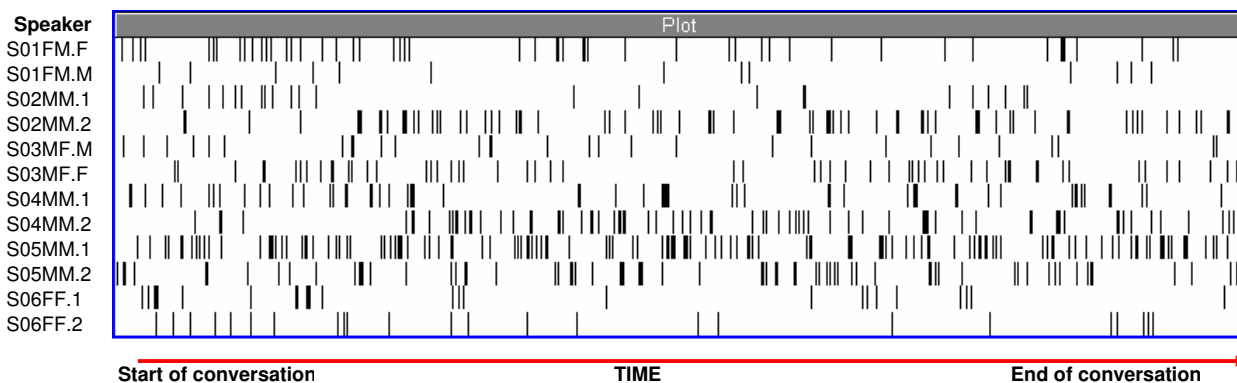
Appendix 6.13: Scatter plots representing the use of spoken and non-verbal backchannels across each video (and speaker) in the five-hour corpus.

Plot 13
A scatter plot of spoken backchannels used with conurrent type B backchanneling nods.

Plot 14
A scatter plot of spoken backchannels used with conurrent type C backchanneling nods.

Plot 15
A scatter plot of spoken backchannels used with conurrent type D backchanneling nods.

Plot 16
A scatter plot of spoken backchannels used with conurrent type E backchanneling nods.

Details of behaviour

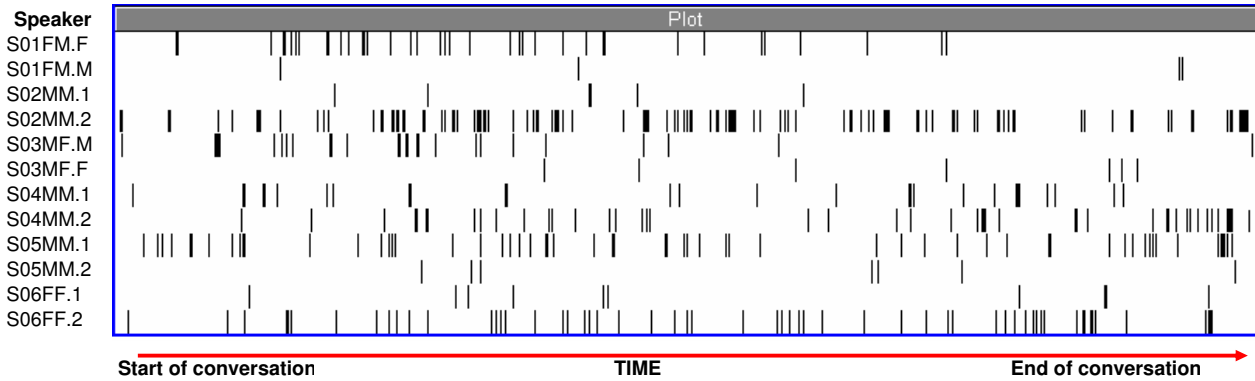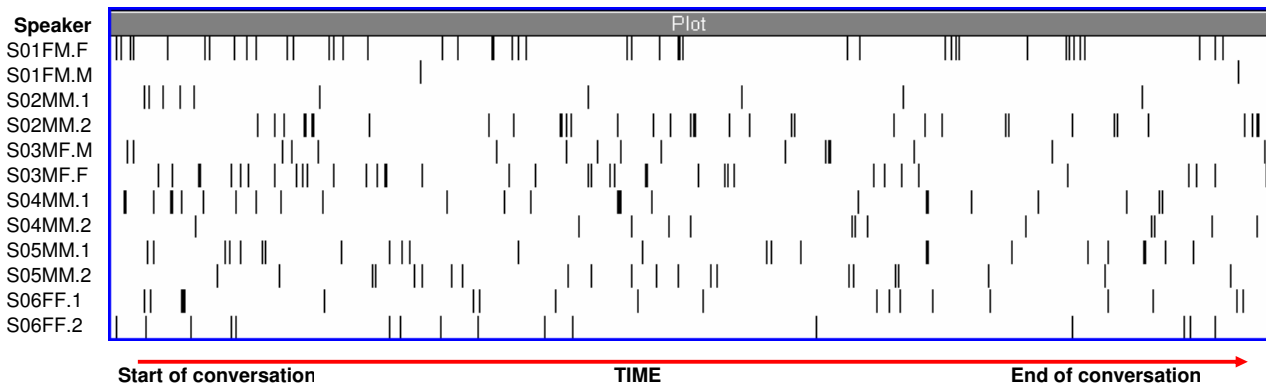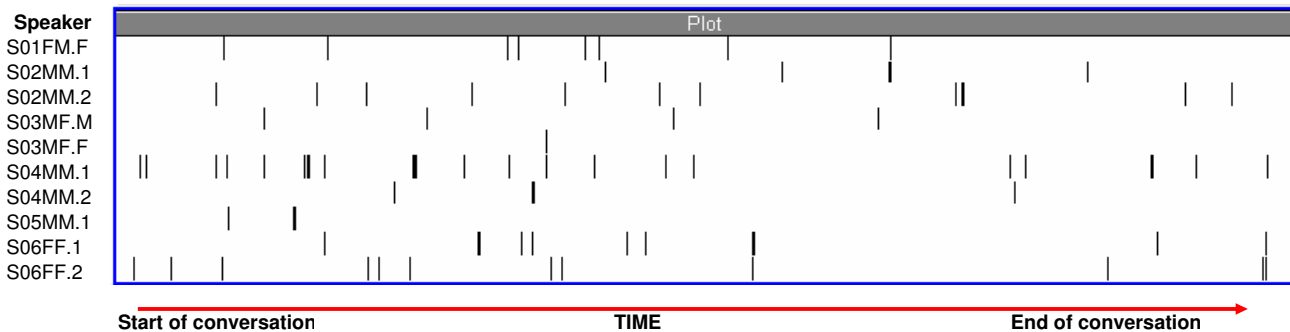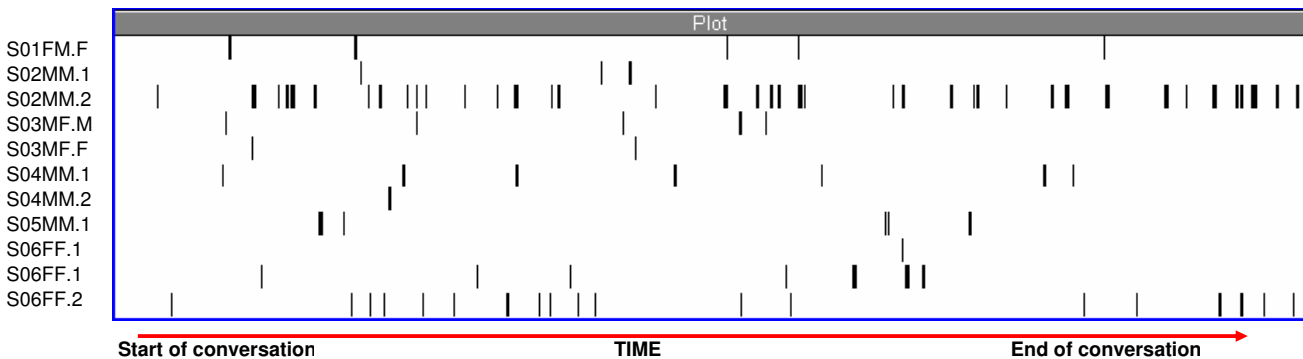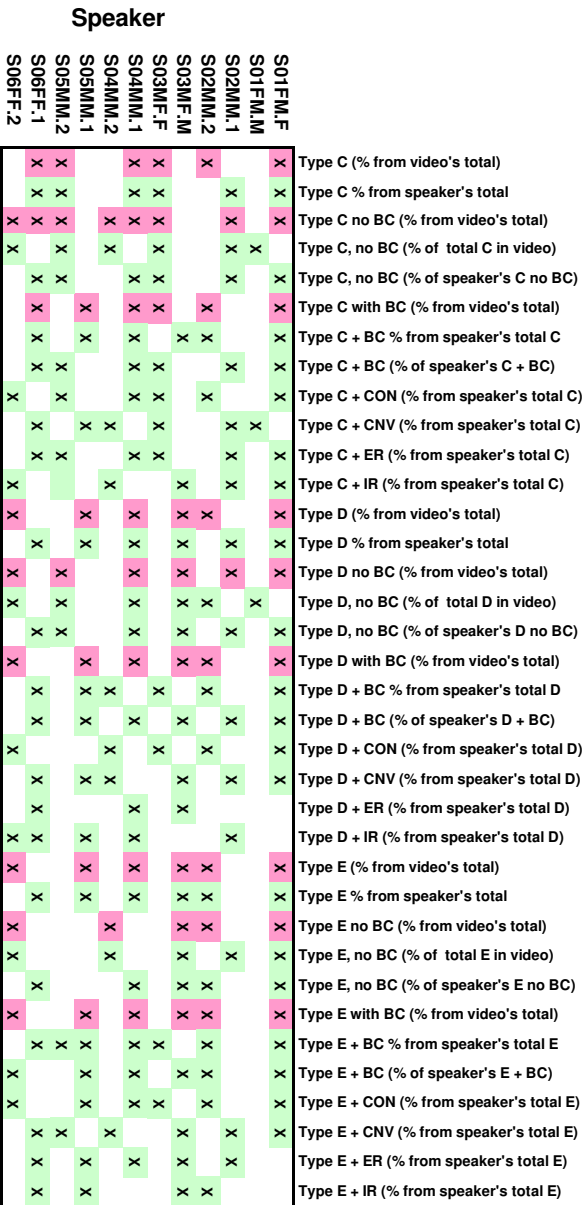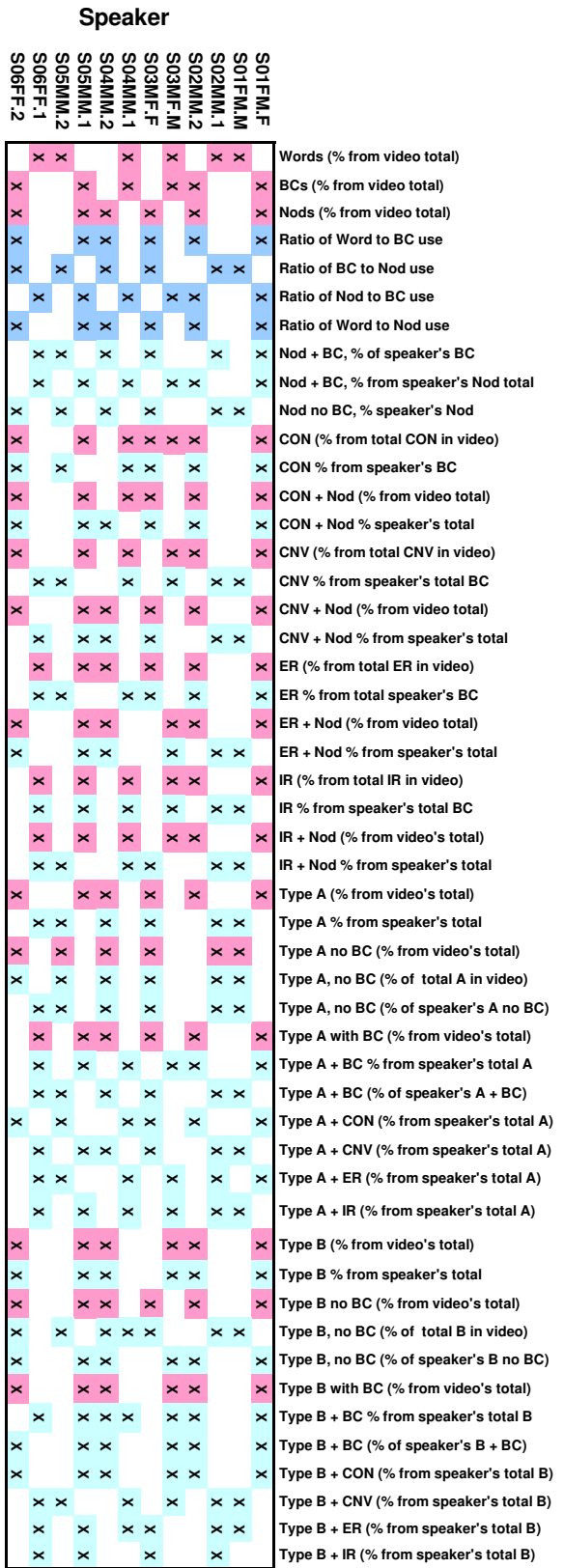|	S06FF.2	S06FF.1	S05MM.2	S05MM.1	S04MM.2	S04MM.1	S03MF.F	S03MF.M	S02MM.2	S02MM.1	S01FM.M	S01FM.F
Words (% from video total)	25	75	62	38	42	58	31	69	40	54	46	87
BCs (% from video total)	64	36	70	47	53	45	67	33	32	13	13	87
Nods (% from video total)	74	26	40	56	44	67	68	68	32	75	68	68
Ratio of Word to BC use	14	78	48	13	14	18	36	7.5	75	121	15	
Ratio of BC to Nod use	14	1	1	1	1	0	1	1	0	3	1	
Ratio of Nod to BC use	2	1	1	1	1	2	1	0	3	1		
Ratio of Word to Nod use	8	67	38	13	14	25	8	10	32	45	18	
Nod + BC, % of speaker's BC	69	74	76	61	58	49	80	52	56	56	54	58
Nod + BC, % from speaker's Nod total	37	64	64	81	57	69	34	54	75	30	20	68
Nod no BC, % speaker's Nod	63	36	41	19	43	31	66	46	25	80	80	32
CON (% from total CON in video)	85	15	33	46	46	54	50	50	96	8	8	92
CON % from speaker's BC	52	16	67	58	58	55	39	32	74	38	38	65
CON + Nod (% from video total)	88	12	32	48	52	57	43	97	3	8	92	
CON + Nod % speaker's total	80	65	74	77	64	75	59	83	50	53		
CNV (% from total CNV in video)	76	24	36	41	59	38	62	95	8.3	92		
CNV % from speaker's total BC	20	35	36	23	30	25	17	23	50	47		
CNV + Nod (% from video total)	52	48	24	53	47	55	45	74	21	79		
CNV + Nod % from speaker's total	33	53	35	41	31	46	31	47	49	27		
ER (% from total ER in video)	42.2	57.8	25.2	74.8	47.8	59.1	40.9	76.8	15.7	84.3		
ER % from total speaker's BC	62	93	74	59	62	85	82	70	61	87		
ER + Nod (% from video total)	85	15	21	79	55	45	39	61	37	44	56	
ER + Nod % from speaker's total	38	7.1	21	41	38	15	28	18	39	13		
IR (% from total IR in video)	36	64	27	14	86	32	68	71	18	82		
IR % from speaker's total BC	4	41	6	3	15	15	33	7	7	14		
IR + Nod (% from video's total)	33	67	29	11	89	42	58	60	25	75		
IR + Nod % from speaker's total	75	86	83	75	50	67	63	50	80	50	33	
Type A (% from video's total)	57	43	44	64	36	75	25	57	57	48	52	
Type A % from speaker's total	17	36	73	57	41	65	43	36	61	31		
Type A no BC (% from video's total)	71	29	35	65	25	81	19	37	75	25		
Type A, no BC (% of total A in video)	60	32	43	42	24	72	53	33	77	25		
Type A, no BC (% of speaker's A no BC)	16	33	77	56	56	71	49	47	59	24		
Type A with BC (% from video's total)	43	57	35	65	43	65	35	47	23	78		
Type A + BC % from speaker's total A	39.7	68.2	56.9	81.8	75.6	28.1	47	22	23	75.4		
Type A + BC (% of speaker's A + BC)	18	68	71	62	69	45	54	37	70	34		
Type A + CON (% from speaker's total A)	52	23	73	63	63	40	76	24	36	59		
Type A + CNV (% from speaker's total A)	30	47	22	31	45	47	42	21	48	37		
Type A + ER (% from speaker's total A)	0	3		4	10	4	6	16	2			
Type A + IR (% from speaker's total A)	17	27	1.3	2.1	1.2	4.8	8.8	29	12	7.1	2	
Type B (% from video's total)	91	9.3	13	87	35	43	57	88	30	70		
Type B % from speaker's total	62	18	6	25	20	10	26	11	28	31		
Type B no BC (% from video's total)	92	8	23	77	35	56	44	46	80	54		
Type B, no BC (% of total B in video)	72	59	45	45	73	43	26	86	42			
Type B, no BC (% of speaker's B no BC)	71	30	32	31	11	24	39	30	42			
Type B with BC (% from video's total)	87	13	10	65	11	95	74	10	90			
Type B + BC % from speaker's total B	28	41	55	76	55	27	74	36	58			
Type B + BC (% of speaker's B + BC)	46	12	5	24	8	25	39	82	26			
Type B + CON (% from speaker's total B)	66	0	67	68	58	73	82	16	50	76		
Type B + CNV (% from speaker's total B)	25	60	33	28	45	3	24					
Type B + ER (% from speaker's total B)	5	20	0	17	33	25	0					
Type B + IR (% from speaker's total B)	3	20	20	1	0	0	13	12				

Details of behaviour

	S06FF.2	S06FF.1	S05MM.2	S05MM.1	S04MM.2	S04MM.1	S03MF.M	S02MM.2	S02MM.1	S01FM.M	S01FM.F
Type C (% from video's total)	46	54	54	46	41	59	30	51	49	15	85
Type C % from speaker's total	8	26	20	11	10	24	21	12	25	10	27
Type C no BC (% from video's total)	50	50	75	25	50	75	12	12	88	40	60
Type C, no BC (% of total C in video)	41	34	32	13	38	27	38	9.5	73	80	21
Type C, no BC (% of speaker's C no BC)	5	26	16	9	18	17	26	5	10	4	18
Type C with BC (% from video's total)	43	57	48	37	37	63	34	78	22	4	96
Type C + BC % from speaker's total C	59	66	68	87	62	63	51	90	20	79	
Type C + BC (% of speaker's C + BC)	12	27	23	12	11	36	24	15	10	31	
Type C + CON (% from speaker's total C)	38	5	44	26	38	29	25	51	0	41	
Type C + CNV (% from speaker's total C)	31	62	40	67	30	25	37	45	100	50	
Type C + ER (% from speaker's total C)	0	5	8	4	0	7	3	10	0	0	
Type C + IR (% from speaker's total C)	31	29	8	4	25	1	11	2	36	0	0
Type D (% from video's total)	65	35	33	67	8	11	11	89	61	6	94
Type D % from speaker's total	7		11	1	1	16	5	3	4	1	8
Type D no BC (% from video's total)	80	20	100	0	0	100	0	100	10	90	
Type D, no BC (% of total D in video)	53	23	55	0	44	36	29	50	38	14	210
Type D, no BC (% of speaker's D no BC)	6	7	1	0	0	19	0	6	1	13	
Type D with BC (% from video's total)	55	45	0	100	13	88	0	80	65	35	100
Type D + BC % from speaker's total D	50	76.9	0	100	100	63.6	0	100	85.7	43.8	
Type D + BC (% of speaker's D + BC)	9	13	N	1	2	15	1	4	12		
Type D + CON (% from speaker's total D)	62	23	N	50	41	100	25	94	50	N	38
Type D + CNV (% from speaker's total D)	23	54	N	67	43	50	6.3	17	N	63	
Type D + ER (% from speaker's total D)	0	8	N	0	5	0	25	0	N	0	
Type D + IR (% from speaker's total D)	15	15	N	33	11	0	0	33	N	0	
Type E (% from video's total)	71	29	17	83	31	69	27	73	89	11	100
Type E % from speaker's total	7	8	1	2	4	1	3	12	3	0	4
Type E no BC (% from video's total)	71	29	N	N	60	40	25	75	82	18	100
Type E, no BC (% of total E in video)	21	20	0	0	75	22	38	21	40	N	25
Type E, no BC (% of speaker's E no BC)	2	5	0	0	23	3	4	10	2	0	3
Type E with BC (% from video's total)	70	30	17	83	13	88	71	92	8.3	100	
Type E + BC % from speaker's total E	79	80	100	100	25	78	67	63	79	60	N	75
Type E + BC (% of speaker's E + BC)	15	59.1	0	2	12.5	1	5	6	13	6	4	55.6
Type E + CON (% from speaker's total E)	15	59.1	0	12.5	55.6	100	44.4	72.9	N	55.6	
Type E + CNV (% from speaker's total E)	36	80	100	50	39	11	20	75	N	44	
Type E + ER (% from speaker's total E)	0	7	0	25	6	0	0	33	4	25	
Type E + IR (% from speaker's total E)	5	13	0	13	0	0	11	3	N	N	0

KEY

Colour	Meaning
(pink/red)	% of use from total across both speakers
(light blue)	% of use from individual speaker total
(blue)	Ratio of first parameter: second parameter — e.g. Nods: Words

Symbol	Meaning
no	Without a concurrent backchannel (of another type)
CON	Spoken backchannel function (CON, CNV, ER or IR)
N	Negligible result (result under 1%)
BC	Spoken Backchannels
+	Concurrent spoken and non-verbal backchannels
X	Denotes the highest % of use across speakers in a given video, according to parameters listed
Type A	Head nod type (from A to E)

Appendix 6.14: Exploring the relationships between the use of spoken and non-verbal backchannels and their discursive functions (across the five-hour corpus).

Appendix 6.15: Mapping the patterns between the use of spoken and non-verbal backchannels and their associated discursive functions (across the five-hour corpus).

Details of behaviour

Speaker (columns): S06FF.2, S06FF.1, S05MM.2, S05MM.1, S04MM.2, S04MM.1, S03MF.F, S03MF.M, S02MM.2, S02MM.1, S01FM.M, S01FM.F

Rows (Details of behaviour):

- Words (% from video total)
- BCs (% from video total)
- Nods (% from video total)
- Ratio of Word to BC use
- Ratio of BC to Nod use
- Ratio of Nod to BC use
- Ratio of Word to Nod use
- Nod + BC, % of speaker's BC
- Nod + BC, % from speaker's Nod total
- Nod no BC, % speaker's Nod
- CON (% from total CON in video)
- CON % from speaker's BC
- CON + Nod (% from video total)
- CON + Nod % speaker's total
- CNV (% from total CNV in video)
- CNV % from speaker's total BC
- CNV + Nod (% from video total)
- CNV + Nod % from speaker's total
- ER (% from total ER in video)
- ER % from total speaker's BC
- ER + Nod (% from video total)
- ER + Nod % from speaker's total
- IR (% from total IR in video)
- IR % from speaker's total BC
- IR + Nod (% from video's total)
- IR + Nod % from speaker's total
- Type A (% from video's total)
- Type A % from speaker's total
- Type A no BC (% from video's total)
- Type A, no BC (% of total A in video)
- Type A, no BC (% of speaker's A no BC)
- Type A with BC (% from video's total)
- Type A + BC % from speaker's total A
- Type A + BC (% of speaker's A + BC)
- Type A + CON (% from speaker's total A)
- Type A + CNV (% from speaker's total A)
- Type A + ER (% from speaker's total A)
- Type A + IR (% from speaker's total A)
- Type B (% from video's total)
- Type B % from speaker's total
- Type B no BC (% from video's total)
- Type B, no BC (% of total B in video)
- Type B, no BC (% of speaker's B no BC)
- Type B with BC (% from video's total)
- Type B + BC % from speaker's total B
- Type B + BC (% of speaker's B + BC)
- Type B + CON (% from speaker's total B)
- Type B + CNV (% from speaker's total B)
- Type B + ER (% from speaker's total B)
- Type B + IR (% from speaker's total B)

Details of behaviour

Speaker (columns): S06FF.2, S06FF.1, S05MM.2, S05MM.1, S04MM.2, S04MM.1, S03MF.F, S03MF.M, S02MM.2, S02MM.1, S01FM.M, S01FM.F

Rows (Details of behaviour):

- Type C (% from video's total)
- Type C % from speaker's total
- Type C no BC (% from video's total)
- Type C, no BC (% of total C in video)
- Type C, no BC (% of speaker's C no BC)
- Type C with BC (% from video's total)
- Type C + BC % from speaker's total C
- Type C + BC (% of speaker's C + BC)
- Type C + CON (% from speaker's total C)
- Type C + CNV (% from speaker's total C)
- Type C + ER (% from speaker's total C)
- Type C + IR (% from speaker's total C)
- Type D (% from video's total)
- Type D % from speaker's total
- Type D no BC (% from video's total)
- Type D, no BC (% of total D in video)
- Type D, no BC (% of speaker's D no BC)
- Type D with BC (% from video's total)
- Type D + BC % from speaker's total D
- Type D + BC (% of speaker's D + BC)
- Type D + CON (% from speaker's total D)
- Type D + CNV (% from speaker's total D)
- Type D + ER (% from speaker's total D)
- Type D + IR (% from speaker's total D)
- Type E (% from video's total)
- Type E % from speaker's total
- Type E no BC (% from video's total)
- Type E, no BC (% of total E in video)
- Type E, no BC (% of speaker's E no BC)
- Type E with BC (% from video's total)
- Type E + BC % from speaker's total E
- Type E + BC (% of speaker's E + BC)
- Type E + CON (% from speaker's total E)
- Type E + CNV (% from speaker's total E)
- Type E + ER (% from speaker's total E)
- Type E + IR (% from speaker's total E)

KEY

- % of use from total across both speakers
- % of use from individual speaker total
- Ratio of first parameter: second parameter
  - e.g. Nods : Words
- **no** — Without a concurrent backchannel (of another type)
- **CON** — Spoken backchannel function (CON, CNV, ER or IR)
- **N** — Negligible result (result under 1%)
- **BC** — Spoken Backchannels
- **+** — Concurrent spoken and non-verbal backchannels
- **X** — Denotes the highest % of use across speakers in a given video, according to parameters listed
- **Type A** — Head nod type (from A to E)

Appendix 6.16: Charting the lexical clusters that are most frequently exist in the immediate co-text of non-verbal backchannel use (across the five-hour corpus).

Rank	Cluster	Freq.	Rank	Cluster	Freq.	Rank	Cluster	Freq.
1	nn yeah i	13	41	look at #nn	4	79	nn corpus linguistics	3
2	erm pause #nn	8	42	nn yeah so	4	80	nn i mean	3
3	you know what	8	43	some kind of	4	81	nn in the	3
4	do you know	7	44	some of the	4	82	nn is interesting	3
5	kind of #nn	7	45	talking about #nn	4			
6	of the #nn	7	46	to look at	4	83	nn one of	3
7	going to #nn	7	47	yeah i think	4	84	nn the historical	3
8	i think #nn	7	48	you need to	4			
9	nn do you	7	49	a chapter #nn	3	85	nn well that's	3
10	nn i think	7	50	a sort of	3	86	nn yeah well	3
11	nn yeah and	7	51	and that #nn	3	87	on the #nn	3
12	see what i	7	52	as a #nn	3	88	pause #nn the	3
13	sort of #nn	7	53	at that particular	3	89	pause erm pause	3
14	the idea of	7						
15	and i think	5	54	ba# yeah yeah	3	90	some of them	3
16	and then you	5	55	but i mean	3	91	space theory #nn	3
17	in terms of	5	56	change in the	3			
18	in the #nn	5	57	conceptual #nn metaphor	3	92	study is it	3
19	know what i	5				93	terms of the	3
20	nn that's right	5	58	do you see	3	94	that kind of	3
21	nn yeah yeah	5	59	er and #nn	3	95	that sort of	3
22	nn you know	5	60	erm the #nn	3	96	that would be	3
23	of #nn the	5	61	i could #nn	3	97	the #nn the	3
24	one of the	5	62	i mean you	3	98	the hotel #nn	3
25	quite a lot	5	63	i think i	3	99	the hotel and	3
26	to do with	5	64	i think it's	3	100	the kind of	3
27	what i mean	5	65	i want to	3	101	the the #nn	3
28	yeah #nn yeah	5	66	idea of erm	3	102	to have to	3
29	you can #nn	5	67	in a sense	3	103	to somebody else	3
30	you know #nn	5	68	in order to	3			
31	you see what	5	69	is a #nn	3	104	to the #nn	3
32	you want to	5	70	is is #nn	3	105	want to #nn	3
33	a kind of	4	71	is that the	3	106	way of #nn	3
34	a lot of	4	72	it to you	3	107	well i think	3
35	a way of	4	73	kind of the	3	108	what #nn i	3
36	and and #nn	4	74	looking at #nn	3	109	yeah i mean	3
37	going to have	4	75	needs to be	3	110	you know just	3
38	have to #nn	4	76	nn and the	3	111	you know the	3
39	i mean #nn	4	77	nn as a	3	112	you're going to	3
40	in a #nn	4	78	nn at the	3			

KEY:

nn = Indicates the position of non-verbal backchannel behaviour

#nn = Indicates the position of non-verbal backchannel behaviour

Freq. = Frequency of use

Appendix 6.17: Charting the lexical clusters that most frequently exist in the immediate co-text of spoken backchannels without concurrent head nods (across the five-hour corpus).

Rank	Cluster	Freq.
1	pause bn#	15
2	bn# and i	10
3	bn# erm and	8
4	bn# you know 7	7
5	bn# erm i	6
6	bn# erm pause	6
7	bn# in terms	6
8	in terms of	6
9	ba# bn# yeah	5
10	bn# and also	5
11	bn# and the	5
12	bn# and then	5
13	bn# and you	5
14	bn# erm but	5
15	erm i think	5
16	erm pause and	5
17	you know bn#	5
18	ba# bn# ba#	4
19	bn# #nn yeah	4
20	bn# and that	4
21	bn# bn# erm	4
22	bn# bn# yeah	4
23	bn# erm so	4
24	bn# i mean	4
25	bn# i think	4
26	bn# in a	4
27	bn# so i	4
28	bn# that you	4
29	bn# um and	4
30	bn# which is	4
31	bn# yeah yeah	4
32	erm bn# and	4
33	it bn# and	4
34	it would be	4
35	to do bn#	4
36	what i mean	4
37	you want to	4
38	and i just	3
39	and you know	3

Rank	Cluster	Freq.
40	as it were	3
41	as well because	3
42	at the same	3
43	away bn# interesting	3
44	ba# bn# so	3
45	bn# and as	3
46	bn# and er	3
47	bn# and erm	3
48	bn# as you	3
49	bn# at the	3
50	bn# but also	3
51	bn# but you	3
52	bn# discourse bn#	3
53	bn# do you	3
54	bn# erm bn#	3
55	bn# erm the	3
56	bn# like that	3
57	bn# so you	3
58	bn# that i	3
59	discourse bn# erm	3
60	do you see	3
61	i don't know	3
62	i think bn#	3
63	i think i	3
64	in relation to	3
65	looking at the	3
66	nn yeah yeah	3
67	of it bn#	3
68	on the website	3
69	pause erm pause	3
70	see what i	3
71	straight away bn#	3
72	that i can	3
73	that isn't it	3
74	the bn# the	3
75	there's what we	3
76	things bn#	3

KEY:

ba = Indicates the position of spoken backchannel behaviour (without nods)

bn = Indicates the position of spoken backchannel behaviour (with nods)

Freq. = Frequency of use

Appendix 6.18: Charting the lexical clusters that most frequently exist in the immediate co-text of spoken backchannels with concurrent head nods (across the five-hour corpus).

Rank	Cluster	Freq.	Rank	Cluster	Freq.	Rank	Cluster	Freq.
1	pause ba#	67	48	you know what	6	95	ba# um and	4
2	ba# you know	37	49	a little bit	5	96	i think you	4
3	ba# erm and	30	50	and ba# and	5	97	if ba# that	4
4	in terms of	21	51	and sort of	5	98	if you like	4
5	what i mean	19	52	ba# and er	5	99	in order to	4
6	i mean ba#	18	53	ba# because ba#	5	100	it ba# and	4
7	ba# do you	17	54	ba# bn# yeah	5	101	make sense ba#	4
8	ba# i think	17	55	ba# but it's	5	102	mental health ba#	4
9	ba# and the	16	56	ba# erm i	5	103	pause you know	4
10	ba# and erm	13	57	ba# of the	5	104	postcards ba# and	4
11	ba# erm but	13	58	ba# so it	5	105	that ba#	4
12	ba# and i	12	59	ba# so that	5	106	the ba#	4
13	ba# and then	12	60	ba# this is	5	107	the hotel ba#	4
14	see what i	12	61	ba# where you	5	108	you know it's	4
15	ba# and and	10	62	does that make	5	109	a ba# positive	3
16	ba# and that	10	63	erm and so	5	110	a lot of	3
17	ba# erm so	10	64	in relation to	5	111	a memory ba#	3
18	ba# in terms	10	65	it ba# erm	5	112	about ba# the	3
19	you know ba#	10	66	of it ba#	5	113	an hour ba#	3
20	you see what	10	67	one of the	5	114	and ba#	3
21	ba# and so	9	68	space theory ba#	5	115	and ba# erm	3
22	ba# erm pause	9	69	that make sense	5	116	and er pause	3
23	ba# so i	9	70	a bit more	4	117	and erm pause	3
24	do you see	9	71	as opposed to	4	118	and i was	3
25	and i think	8	72	a well ba#	4	119	and so i	3
26	ba# pause	8	73	a ground level	4	120	and so that's	3
27	ba# as well	7	74	ba# and #nn	4	121	and the idea	3
28	ba# but i	7	75	ba# and if	4	122	are going to	3
29	ba# cos i	7	76	ba# and just	4	123	as if ba#	3
30	ba# i mean	7	77	ba# and sort	4	124	as it were	3
31	ba# in the	7	78	ba# and you're	4	125	at the moment	3
32	ba# kind of	7	79	ba# as i	4	126	ba# a little	3
33	ba# rather than	7	80	ba# because it	4	127	ba# a lot	3
34	ba# which is	7	81	ba# bn# so	4	128	ba# about what	3
35	isn't it ba#	7	82	ba# but it	4	129	ba# and a	3
36	you need to	7	83	ba# but then	4	130	ba# and also	3
37	ba# and it	6	84	ba# does that	4	131	ba# and how	3
38	ba# and that's	6	85	ba# er so	4	132	ba# and i've	3
39	ba# bn# ba#	6	86	ba# erm because	4	133	ba# and it's	3
40	ba# but also	6	87	ba# erm yeah	4	134	ba# as it	3
41	ba# but ba#	6	88	ba# erm you	4	135	ba# as opposed	3
42	ba# if you	6	89	ba# it's a	4	136	ba# ba# ba#	3
43	ba# so you	6	90	ba# or are	4	137	ba# ba# erm	3
44	i think ba#	6	91	ba# so if	4	138	ba# ba# theoretical	3
45	know what i	6	92	ba# so it's	4	139	ba# bn# erm	3
46	mean ba# and	6	93	ba# so that's	4	140	ba# but in	3
47	that kind of	6	94	ba# sort of	4	141	ba# but yeah	3

Appendix 6.18: Charting the lexical clusters that most frequently exist in the immediate co-text of spoken backchannels with concurrent head nods (across the complete corpus).

Rank	Cluster	Freq.	Rank	Cluster	Freq.	Rank	Cluster	Freq.
142	ba# discourse and	3	182	cos i think	3	224	particular ba# because	3
143	ba# er and	3	183	critical discourse analysis	3	225	particular experience ba#	3
144	ba# er or	3	184	do it ba#	3	226	perhaps ba# because	3
145	ba# erm bn#	3	185	do you know	3	227	possibly a ba#	3
146	ba# erm to	3	186	does that sound	3	228	property er ba#	3
147	ba# erm which	3	187	erm and erm	3	229	serves to intensify	3
148	ba# evaluative property	3	188	erm and i	3	230	should ba#	3
149	ba# from ba#	3	189	erm ba# and	3	231	so i think	3
150	ba# how you're	3	190	erm ba# but	3	232	so if i	3
151	ba# i thought	3	191	erm ba# so	3	233	so you need	3
152	ba# in effect	3	192	for a particular	3	234	somebody's head ba#	3
153	ba# in order	3	193	for example ba#	3	235	something ba# in	3
154	ba# in relation	3	194	for it ba#	3	236	something like that	3
155	ba# is that	3	195	from ba# this	3	237	sort of erm	3
156	ba# it would	3	196	going to do	3	238	sort of thing	3
157	ba# just to	3	197	ground level ba#	3	239	spaces ba# and	3
158	ba# of of	3	198	head ba# ba#	3	240	terms ba# and	3
159	ba# on a	3	199	i can imagine	3	241	terms of your	3
160	ba# one of	3	200	i think it's	3	242	the detective ba#	3
161	ba# or whether	3	201	i think that's	3	243	the findings ba#	3
162	ba# over time	3	202	in a sense	3	244	the landscape ba#	3
163	ba# perhaps ba#	3	203	in particular ba#	3	245	the nineteenth century	3
164	ba# so what	3	204	in the text	3	246	the space theory	3
165	ba# spoken narrative	3	205	is that alright	3	247	the text ba#	3
166	ba# that i	3	206	it ba#	3	248	the theatre space	3
167	ba# that you've	3	207	it ba# i	3	249	theoretical side ba#	3
168	ba# theoretical side	3	208	it ba# so	3	250	to do ba#	3
169	ba# this ba#	3	209	it would be	3	251	to do with	3
170	ba# to the	3	210	know ba# it	3	252	to explain ba#	3
171	ba# type of	3	211	like that ba#	3	253	to intensify ba#	3
172	ba# what you	3	212	listening post ba#	3	254	to look at	3
173	ba# when you	3	213	literature ba# review	3	255	top of the	3
174	ba# where there	3	214	look at the	3	256	topics ba# and	3
175	ba# you might	3	215	memory ba# yeah	3	257	well ba#	3
176	ba# you need	3	216	miles an hour	3	258	what sort of	3
177	ba# your work	3	217	of ba# building	3	259	you know you	3
178	because ba# i	3	218	of my own	3			
179	bowen and greene	3	219	on mental health	3			
180	but in terms	3	220	one ba# and	3			
181	cheeky ba# uh	3	221	or ba#	3			
			222	or are you	3			
			223	or something ba#	3			

KEY: **ba** = Indicates the position of spoken backchannel behaviour (without nods)

bn = Indicates the position of spoken backchannel behaviour (with nods)

Freq. = Frequency of use

References

Abercrombie, D. (1963) *Studies in phonetics and linguistics.* London: Oxford University Press.

Adolphs, S. (2006) *Introducing electronic text analysis- A practical guide for language and literary studies.* London: Routledge.

Adolphs, S. (2008*) Corpus and Context: investigating pragmatic functions in spoken discourse.* Amsterdam: John Benjamins.

Aijmer, K. (1987) Oh and Ah in English conversation. In Meijs, W. (Ed.) *Corpus Linguistics and Beyond: Proceedings of the Seventh International Conference on English Language Research on Computerized Corpora.* Amsterdam: Rodopi. pp.61-86.

Aist, G., Allen, J., Campana, E., Galescu, L., G´omez Gallo, C., Stoness, S., Swift, M. and Tanenhaus, M. (2006) Software architectures for incremental understanding of human speech. *Proceedings of Interspeech 2006.* Pittsburgh PA: USA. pp.1922-1925.

Alibali, M. W., Kita, S., and Young, A. (2000) Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes, 15*: pp.593-613.

Alibali, M. W., Heath, D. C., and Myers, H. J. (2001) Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language, 44*: pp.169-188.

Allen, J. and Core, M. (1997) Draft of DAMSL: Dialog Act Markup in Several Layers [online report]. Available at: http://www.cs.rochester.edu/research /speech/damsl/RevisedManual/ [Accessed 10 November 2008].

Allen, D.E. and Guy, R.F. (1974) *Conversation analysis: the sociology of talk.* The Hague: Mouton.

Allwood, J., Björnberg, M., Grönqvist, L., Ahlsen, E. and Ottesjö, C. (2000) The Spoken Language Corpus at the Department of Linguistics, Göteborg University. *FQS- Forum: Qualitative Social Research 1*(3) [online]. Available at: http://www.qualitative-research.net/fqs/ [Accessed 16 December 2008].

Allwood, J., Grönqvist, L., Ahlsén, E. and Gunnarssan, M. (2001) Annotations and tools for an activity based spoken language corpus. *Proceedings of the Second SIGdial Workshop of Discourse and Dialogue 16.* Morristown, NJ: Association for Computational Linguistics. pp.1-10.

Allwood, J., Cerrato, L., Jokinen, K., Navarretta, C and Paggio, P. (2007a) The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena. *Language Resources and Evaluation 41*(3): pp.273-287.

Allwood, J., Kopp, S., Grammer, K., Ahlsén, E., Oberzaucher, E. and Koppensteiner, M. (2007b) The analysis of embodied communicative feedback in multimodal corpora: a prerequisite for behaviour simulation. *Language Resources and Evaluation 41*(3): pp.255–272.

Allwood, J., Nivre, J. and Ahlsén, E. (1993) On the semantics and pragmatics of linguistic feedback. *Journal of semantics 9*(1): pp.1-26.

Altorfer, A., Jossen, S., Würmle, O., Käsermann, M. L., Foppa, K., and Zimmermann H. (2000) Measurement and meaning of head movement behavior in everyday face-to-face communicative interaction. *Behavior Research Methods, Instruments and Computers, 32*(1): pp.17-32.

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S. and Weinert, R. (1991) The HCRC Map Task Corpus. *Language and Speech 34*: pp.351-366.

Argyle, M. (1979) New dimensions in the analysis of social skills. In Wolfgang, A. (Ed.) *Nonverbal behaviour- Applications and cultural implications.* London: Academic Press. pp.19-25.

Argyle, M. (1988) *Bodily Communication.* 2nd ed. London: Methuen.

Ashby, S., Bourban, S., Carletta, J., Flynn, M., Guillemot, M., Hain, T., Kadlec, J., Karaiskos, V., Kraaij, W., Kronenthal, M., Lathoud, G., Lincoln, M., Lisowska, A., McCowan, I., Post, W., Reidsma, D. and Wellner, P. (2005) The AMI Meeting Corpus. *Proceedings of Measuring Behavior 2005.* Wageningen, NL. pp.4-8.

Avilés-Arriaga, H.H. and Sucar, L.E. (2002) Dynamic Bayesian networks for visual recognition of dynamic gestures. *Journal of Intelligent and Fuzzy Systems 12*(3-4): pp.243-250.

Badre, A., Guzdial, M., Hudson, S. and Santos, P. (1995) A user interface evaluation using synchronized video, visualizations and event trace data. *Software Quality Journal, 4*(2): pp.101–113.

Baldry, A. and Thibault, P.J. (2001) Towards Multimodal Corpora. In Aston, G. and Burnard, L. (Eds.) *Corpora in the Description and Teaching of English-Papers from the 5th ESSE Conference.* Bologna: Cooperativa Libraria Universitaria Editrice Bologna. pp.87-102.

Baldry, A. and Thibault, P.J. (2006) *Multimodal Transcription and Text Analysis: A multimedia toolkit and course book.* London: Equinox.

Baroni, M. and Ueyama, M. (2006) Building general and special purpose corpora by Web crawling. *Proceedings of the 13th NIJL International Symposium* [online]. Available at: http://www.tokuteicorpus.jp/result/pdf/2006_004.pdf [Accessed 10 November 2008].

Bateson, G. (1968) Redundancy and Coding. In Sebeok, T.A. (Ed.) *Animal Communication: Techniques of study and results of research.* Bloomington: Indiana University Press. pp.614-626.

Bavelas, J.B. (1994) Gestures as part of speech: methodological implications. *Research on Language and Social Interaction 27*(3): pp.201-221.

Beach, W.A. (1993) Transitional regularities for casual 'okay' usages. *Journal of Pragmatics 19*(44): pp.325-352.

Beach, W.A. (1995) Preserving and constraining options: 'okays' and 'official' priorities in medical interviews. In Morris, G. and Cheneil, R. (Eds.) *Talk of the clinic.* Hillsdale NJ: Lawrence Erlbaum. pp.259-289.

Beattie, G. and Aboudan, R. (1994) Gestures, pauses and speech: An experimental investigation of changing social context on their precise temporal relationships. *Semiotica 99*: pp.239-272.

Beattie, G. and Shovelton, H. (1999) Mapping the range of information contained in the iconic hand gestures that accompany speech. *Journal of Language and Social Psychology 18*: pp.438-463.

Beattie, G. and Shovelton, H. (2002) What properties of talk are associated with the generation of spontaneous iconic hand gestures? *British Journal of Social Psychology 41*(3): pp.403-417.

Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., and Rauzy, S. (2006) Le CID: Corpus of Interactional Data -

protocoles, conventions, annotations. *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix en Provence (TIPA) 25*. pp.25-55.

Bertrand, R., Ferré, G., Blache, P., Espesser, R. and Rauzy, S. (2007) Backchannels revisited from a multimodal perspective. *Proceedings of Auditory-Visual Speech Processing 2007 (AVSP2007) 2007, Kasteel Groenendaal, Hilvarenbeek, The Netherlands*. Available at: http://www.isca-speech.org/archive/avsp07/ av07_P09.html [Accessed 10 November 2008].

Biber, D. (1990) Methodological Issues Regarding Corpus-based Analyses of Linguistic Variation. *Literary and Linguistic Computing* 5 (4): pp.257-69.

Biber, D. (1993) Representativeness in Corpus design. *Literary and Linguistic Computing 8*(4): pp.243-257.

Biber, D., Conrad, S. and Reppen, R. (1998) *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge: Cambridge University Press.

Bilous, F.R. and Krauss, R.M. (1988) Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads. *Language and Communication 8*: pp.183-194.

Bird, S. and Liberman, M. (2001) A formal framework for linguistic annotation. *Speech Communication 33*(1-2): pp.23-60.

Bird, S. and Simons, G. (2000) *White Paper on Establishing and Infrastructure for Open Language Archiving*. Available at: http://www.language-archives.org/docs/ white-paper.html [Accessed 16 December 2008].

Birdwhistell, R.L. (1952) *Introduction to Kinesics: An annotated system for the analysis of body motion and gesture.* Louisville, Kentucky: University of Louisville Press.

Birdwhistell, R.L. (1970) *Kinesics in Context*. University of Pennsylvania Press, Philadelphia.

Blache, P., Bertrand, R. and Ferré, G. (2008) Creating and exploiting multimodal annotated corpora. *Proceedings of Sixth International Conference on Language Resources and Evaluation (LREC) 2008* [online]. pp.110-115. Available at: http://www.lrec-conf.org/proceedings/lr ec2008/pdf/132_paper.pdf [Accessed 16 December 2008].

Black, D.W. (1984) Laughter. *Journal of American Medical Association 252*: pp.2995-2998.

Black, M.J. and Yacoob, Y. (1998) Recognizing facial expression in image sequences using local parameterised modes of image motion. *International Journal on Computer Vision 25*(1): pp.23-48.

Blum-Kulka, S. and Olshtain, E. (1984) Requests and apologies: A cross-cultural study of speech act realization patterns (CCSARP). *Applied Linguistics 5*(3): pp.196-212.

Boas, F. (1940) *Race, Language and Culture*. New York: Macmillan.

Boersma, P. and Weenink, D. (2005) *Praat: doing phonetics by computer (Version 4.3.14)* [computer program]. Available at: http://www.praat.org/ [Accessed 20 March 2006].

Bolinger, D. (1986) *Intonation and its Parts*. Palo Alto: Stanford University Press.

Bongers, H. (1947) *The history and principles of vocabulary control.* Woerden, Holland: Wocopi.

Bourke, V.J. (1962) Rationalism. In Runes, D.D. (Ed.) *Dictionary of Philosophy*. Totowa, NJ: Littlefield, Adams, and Company. p.263.

Brown, P., and Levinson, S. (1987) *Politeness: Some universals in language use*. Cambridge: Cambridge University Press.

Brown, G. and Yule, G. (1983) *Discourse analysis*. Cambridge: Cambridge University Press.

Brown, R. (1986) *Social Psychology.* 2nd ed. New York: Free Press.

Brugman, H and Russel, A. (2004) Annotating multi-media / multi-modal resources with elan. In Lino, M., Xavier, M., Ferreire, F., Costa, R. and Silva, R. (Eds.) *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC) 2004*. Lisbon: Portugal. pp.2065–2068.

Brundell, P. and Knight, D. (2005) *Current Research and Tools to Support Data Intensive Analysis for Digital Records in eSocial Science* [unpublished NCeSS node project report]. The University of Nottingham, England.

Bublitz, W. (1988) *Supportive fellow-speakers and cooperative conversations*. Amsterdam: John Benjamins.

Buck, R. (1990) Using FACS versus communication scores to measure the spontaneous facial expression of emotion in brain-damaged patients. *Cortex 26*: pp.275-280.

Burnard, L. (2005) Developing linguistic corpora: Metadata for corpus work. In Wynne, M. (Ed.) *Developing Linguistic Corpora: a Guide to Good Practice.* Oxford: Oxbow Books. pp.30-46. Available online at: http://ahds.ac.uk/linguistic-corpora/ [Accessed 2 October 2006].

Burns, A. and Siedlhofer, B. (2002) Speaking and Pronunciation. In Schmitt, N. (Ed.) *An Introduction to Applied Linguistics*. London: Arnold. pp.211-232.

Cameron, D. (2001) *Working with spoken discourse.* London: Sage.

Canepari, L. (2005) *A handbook of phonetics.* Munich: Lincom Europa.

Carletta, J., Evert, S., Heid, U., Kilgour, J., Robertson, J., and Voormann, H. (2003) The NITE XML Toolkit: flexible annotation for multi-modal language data. *Behavior Research Methods, Instruments, and Computers, special issue on Measuring Behavior, 35*(3): pp.353-363.

Cathcart, N., Carletta, J. and Klein, E. (2003) A shallow model of backchannel continuers in spoken dialogue. *10th Conference of the European Chapter of the Association for Computational Linguistics* [online]. pp.51-58. Available at: http://www.iccs.informatics.ed.ac.uk/~ewan/Papers/Cathcart: 2003:SMB.pdf [Accessed 4 July 2006].

Cerrato, L. (2002) A comparison between feedback strategies in Human-to-Human and Human-Machine communication. *Proceedings of International Conference of Speech and Language Processing (ICSLP)* Denver, Colorado. pp.557-560.

Cerrato, L. (2004) A coding scheme for the annotation of feedback phenomenon in everyday speech. *Proceeding of the LREC'2004 Workshop on Models of Human Behaviour for the Specification and Evaluation of Multimodal Input and Output Interfaces* [online]. pp.25-28. Available at: http://www.speech.kth.se/~loce/ papers/Loredana%20Cerra tospublicationlistfrom2002.htm [Accessed 29 May 2007].

Cerrato, L., and Skhiri, M. (2003) Analysis and measurement of head movements signalling feedback in face-to-face human dialogues. *Proceedings of AVSP 2003* [online]. pp.251-256. Available at:

http://www.speech.kth.se/~loce/papers/Lore
dana%20Cerratospublicationlistfrom2002.htm [Accessed 29 May 2007].

Chawla, P. and Krauss, R. M. (1994) Gesture and speech in spontaneous and rehearsed narratives. *Journal of Experimental Social Psychology 30*: pp.580-601.

Chen, L., Travis-Rose, R., Parrill, F., Han, X., Tu, J., Huang, Z., Harper, M., Quek, F., McNeill, D., Tuttle, R. and Huang, T. (2005) VACE Multimodal Meeting Corpus. Proceedings of the Workshop on Machine Learning for Multimodal Interaction (MLMI) [online]. Available at: http://groups.inf.ed.ac.uk/mlmi05/tech prog/arch/M-B-1.pdf [Accessed 10 November 2008].

Chomsky, N. (1957) *Syntactic Structures.* The Hague: Mouton and Co.

Chomsky, N. (1965) *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1983) Noam Chomsky's views on the psychology of language and thought. In Rieber, R. (Ed.) *Dialogues on the psychology of language and thought*. New York: Plenum. pp.33-46.

Church, K.W. and Hanks, P. (1990) Word association norms, mutual information and lexicography. *Computational Linguistics 16*(1): pp.22-29.

Church, R.B. and Goldin-Meadow, S. (1986) The mismatch between gesture and speech as an index of transitional knowledge. *Cognition 23*(1): pp.43-71.

Clancy, P.M., Thompson, S., Suzuki, R. and Tao, H. (1996) The conversational use of reactive tokens in English, Japanese and Mandarin. *Journal of Pragmatics 26*: pp.355-387.

Clark, H.H. and Carlson, T.B. (1982) Hearers and speech acts. *Language 58*(2): pp.332-373.

Clark, H.H. and Krych, M.A. (2004) Speaking while monitoring addressees for understanding. *Journal of Memory and Language 50*(1): pp.62-81.

Clark, H.H. and Schaefer, E.F. (1989) Contributing to discourse. *Cognitive Science 13*: pp.259-294.

Coates, J. (1986) *Women, men and language: a sociolinguistic account of sex differences in language*. London: Longman.

Colombo, C., Del Bimbo, A., and Valli, A. (2003) Visual capture and understanding of hand pointing actions in a 3-D environment. *IEEE Transactions on Systems, Man, and Cybernetics 33*(4): pp.677-686.

Comaniciu, D., Ramesh, V. and Meer, P. (2003) Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence 25*(5): pp.564-577.

Conrad, S. (2002) Corpus linguistic approaches for discourse analysis. *Annual Review of Applied Linguistics 22*: pp.75-95.

Cutrone, P. (2005) A case study examining backchannels in conversations between Japanese-British dyads. *Multilingua 24*(3): pp.237-274.

Daniel, L.G. (1998) Statistical significance testing: A historical overview of misuse and misrepresentation with implications for the editorial policies of educational journals. *Research in the schools 5*(2): pp.23-32.

Davis, M. (1979) The state of the art: Past and present trends in body movement research. In Wolfgang, A. (Ed.) *Nonverbal behaviour-Applications and cultural implications*. London: Academic Press. pp.51-66.

Davis, J.W. and Vaks, S. (2001) A perceptual user interface for recognizing head gesture acknowledgments. *ACM Workshop on Perceptual User Interfaces, PUI 2001* [online]. Available at: http://www.cs.ucsb.edu/PUI/ PUIWorkshop/PUI-2001/a23.pdf [Accessed 6 June 2007].

De Ruiter, J.P., Rossignol, S.F., Vuurpijl, L., Cunningham, D.W. and Levelt, W.J.M. (2003) SLOT- A research platform for investigating multimodal communication. *Behavioural Research Methods, Instruments and Computers 35*(3): pp.408-419.

Deniz, O., Falcon, A., Mendez, J. and Castrillon, M. (2004) Useful computer vision techniques for human-robot interaction. In *Proceeding of the 1st International Conference on Image Analysis and Recognition*, Porto: Portugal. Available online at: http://mozart.dis.ulpgc.es/Gias/Publications/ iciar04.pdf [Accessed 1 May 2006].

Deutscher, J., Blake, A. and Reid, I. (2000) *Articulated body motion capture by annealed particle filtering. Proceedings of the IEEE Conference on Computer Vision Pattern Recognition, 2000*.

Dittman, A. and Llewellyn, L. (1968) Relationships between vocalizations and head nods as listener responses. *Journal of Personality and Social Psychology 9*: pp.79-84.

Dittman, A.T. (1960) Relationship between body movements and moods in interviews. *Journal of Abnormal and Social Psychology 61*: pp.341-347.

Dixon, J.A. and Foster, D.H. (1998) Gender, social context and backchannel responses. *Journal of Social Psychology 138*(1): pp.134-136.

Dobrogaev, S. M. (1929) Ucnenie o reflekse v problemakh iazykovedeniia [Observations on reflexes and issues in language study]. *Iazykovedenie I Materializm*: pp.105-73.

Drummond, K. and Hopper, R. (1993a) Backchannels revisited: Acknowledgement tokens and speaker incipiency. *Research on Language and Social Interaction 26*(2): pp.157-177.

Drummond, K. and Hopper, R. (1993b) Some uses of Yeah. *Research on Language and Social Interaction 26*(2): pp.203-212.

Du Bois, J.W., Schuetze-Coburn, S., Paolino, D. and Cumming, S. (1992) Discourse transcription. *Santa Barbara Papers in Linguistics*, Volume 4, UC Santa Barbara.

Du Bois, J., Schuetze-Coburn, S., Paolino, D. and Cumming, S. (1993) Outline of discourse transcription. In Edwards, J.A. and Lampert, M.D. (Eds.) *Talking data: Transcription and coding methods for language research*. Hillsdale, NJ: Lawrence Erlbaum. pp.45-89.

Duncan, S. (1972) Some signals and rules for taking speaking turns in conversation. *Journal of personality and social psychology 23*(2): pp.283-292.

Duncan, S. and Niederehe, G. (1974) On signalling that it's your turn to speak. *Journal of experimental social psychology 10*(3): pp.234-47.

Duncan, S. D. and Fiske, D. W. (1977) *Face-to-Face Interaction: Research, Methods, and Theory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Dunning, T. (1993) Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics 19*(1): pp.61-74.

Duranti, A. and Goodwin, C. (Eds.) (1992) *Rethinking Context: Language as an Interactive Phenomenon*. Cambridge University Press.

Dybkjær, L. and Ole Bernsen, N. (2004) Recommendations for natural interactivity and multimodal annotation schemes. *Proceedings of the LREC'2004 Workshop on Multimodal Corpora, Lisbon* [online]. pp.5-8. Available at: http://www.multimodal-corpora.org/lrec04.html [Accessed 1 August 2008].

Edwards, J. (1993) Principles and contrasting systems of discourse transcription. In Edwards, J. and Lampert, M. (Eds.) *Talking Data: Transcription and coding in discourse research*. Hillsdale, NJ: Lawrence Erlbaum Associates. pp.3-44

Efron, D. (1941) *Gesture, Race and Culture.* The Hague: Mouton and Co.

Eggins, S. and D. Slade. (1997) *Analysing casual conversation*. London: Cassell.

Eckert, P. and Rickford, J. (Eds.) (2001) *Style and sociolinguistic variatio*n. New York and Cambridge, Cambridge University Press.

Ekman, P. (1982) *Emotion in the Human Face*. 2nd ed. Cambridge: Cambridge University Press.

Ekman, P. (1997) Conclusion: What we have learned by measuring facial behavior. In Ekman, P. and Rosenberg, E. (Eds.) *What the face reveals*. New York: Oxford University Press. pp.469-485.

Ekman, P. and Friesen, W. (1968) Nonverbal behavior in psychotherapy research. In Shlien, J. (Ed.) *Research in Psychotherapy Volume III*. American Psychological Association. pp.179-216.

Ekman, P. and Friesen, W. (1969) The repertoire of non-verbal behavior: Categories, origins, usage and coding. *Semiotica 1*(1): pp.49-98.

Ekman, P. and Friesen, W. (1978) *FACS- Facial Action Coding System* [online]. Carnegie Mellon School of Computer Science. Available at: http://www.cs.cmu.edu/afs/cs/project/face/www/facs.htm. [Accessed 2006-02-10].

El Kaliouby, R. and Robinson, P. (2004) Real-time inference of complex mental states from facial expressions and head gestures. *Workshop on Real-Time vision for HCI, IEEE conference on computer vision and pattern recognition* [online]. pp.950-953. Available at: http://ieeexplore.ieee.org/iel5/9515/30163/01384952.pdf [Accessed 1 May 2006].

Evans, D. and Naeem, A. (2007) Using visual tracking to link text and gesture in studies of natural discourse. *Proceedings of the Cross Disciplinary Research Group Conference, University of Nottingham* [online]. Available at: http://www.mrl.nott.ac.uk/~axc/DReSS_Outputs/CDRG_2007.pdf [Accessed 9 February 2008].

Evans, J.L., Alibali, M.W. and McNeill, N.M. (2001) Divergence of verbal expression and embodied knowledge: Evidence from speech and gesture in children with specific language impairment. *Language and Cognitive Processes 16*(2-3): pp.309-331.

Feke, M.S. (2003) Effects of native language and sex on back-channel behaviour. In *First Workshop on Spanish Sociolinguistics*, Somerville, MA, 2003. pp.96–106.

Fellegy, A.M. (1995) Patterns and functions of minimal response. *American Speech 70*(2): pp.186-199.

Ferrari, G. (1997) Types of contexts and their role in multimodal communication. *Computational Intelligence 13*(3): pp.414-426.

Fetzer, A. (2004) *Recontextualizing context: grammaticality meets appropriateness.* Amsterdam: Benjamins.

Firth, J. (1957) *Papers in linguistics.* Oxford: Oxford University Press.

Flowerdew, L. (2004) The argument for using English specialised corpora to understand academic and professional language. In Connor, U. and Upton, T.A. (Eds.) *Discourse in the professions- perspectives from Corpus Linguistics.* John Benjamins publishing company, Amsterdam. pp.11-33.

Foster, M.E. and Oberlander, J. (2007) Corpus-based generation of head and eyebrow motion for an embodied conversational agent. *Language Resources and Evaluation 41*(3/4): pp.305–323.

Fraser, B. (1999) What are discourse markers? *Journal of pragmatics 31*: pp.931-952.

French, A., Greenhalgh, C., Crabtree, A., Wright, W., Brundell, B., Hampshire, A. and Rodden, T. (2006) Software Replay Tools for Time-based Social Science Data. *Proceedings of the 2nd annual international e-Social Science conference* [online]. Available at: http://www.ncess.ac.uk/events/ conference/2006/papers/abstracts/FrenchSoftwareReplayTools.shtml [Accessed 16 November 2006].

Frey, S., Hirsbrunner, H.P., Florin, A., Daw, W. and Crawford, R. (1983) A unified approach to the investigation of nonverbal and verbal behaviour in communication research. In Doise, W. and Moscovici, S. (Eds.) *Current issues in European Social Psychology.* Cambridge: Cambridge University Press. pp.143-199.

Fries, C. and Traver, A. (1940) *English word lists: a study of their adaptability and instruction.* Washington, DC: American Council of Education.

Fries, C.C. (1952) *The structure of English.* New York: Harcourt, Brace and company.

Gardner, R. (1997a) The conversation object mm: A weak and variable acknowledging token. *Research on Language and Social Interaction 30*(2): pp.131-156.

Gardner, R. (1997b) The listener and minimal responses in conversational interaction. *Prospect 12*(2): pp.12-32.

Gardner, R. (1998) Between speaking and listening: The vocalisation of understanding. *Applied Linguistics 19*(2): pp.204-224.

Gardner, R. (2001) *When listeners talk: response tokens and listener stance.* London: John Benjamin's.

Garofolo, J., Laprun, C., Michel, M., Stanford, V. and Tabassi, E. (2004) The NIST Meeting Room Pilot Corpus. *Proceedings of the 4th Language Resources and Evaluation Conference (LREC) 2004* [online]. Available at: http://www.nist.gov/ speech/test_beds/meeting_corpus_1/index.html [Accessed 16 December 2008].

Garside, R. (1987) The CLAWS Word-tagging System. In Garside, R., Leech, G. and Sampson, G. (Eds.) *The Computational Analysis of English: A Corpus-based Approach.* London: Longman.

Gibbon, D., Moore, R.K. and Winski, R. (Eds.) (1997) *Handbook of Standards and Resources for Spoken Language Systems.* Berlin: Mouton de Gruyter.

Goffman, E. (1963) *Behavior in Public Places.* New York: Free Press.

Goffman, E. (1974) *Frame Analysis.* Cambridge: Harvard University Press.

Golde, C.M. and Gallagher, H.A. (1999) The challenges of conducting interdisciplinary research in traditional Doctoral programs. *Ecosystems 2*: pp.281-285.

Goldin-Meadow, S. (1999) The role of gesture in communication and thinking. *Trends in cognitive sciences 3*(11): pp.419-429.

Goodwin, C. (1981) *Conversational Organisation: Interaction between Speakers and Hearers*. New York: Academic Press.

Goodwin, C. (1986) Between and within: Alternative sequential treatments of continuers and assessments. *Human Studies 9*(2-3): pp.205-217.

Graddol, D., Cheshire, J. and Swann, J. (1994) *Describing Language*. Buckingham: Open University Press.

Granström, B., House, D. and Swerts, M. (2002) Multimodal feedback cues in human-machine interactions. *Proceeding of the Speech Prosody 2002 conference, Aix-en-Provence: Laboratoire Parole et Langage* [online]. pp.347-350. Available at: http://foap.uvt.nl/publications.html [Accessed 9 March 2006].

Greenhalgh, C., French, A., Tennant, P., Humble, J. and Crabtree, A. (2007) From ReplayTool to Digital Replay System. *Proceedings of the 3rd International Conference on e-Social Science ESRC/ NSF* [online]. Available at: http://ess.si.umich.edu/papers/paper161.pdf [Accessed 4 August 2007].

Grice, P. (1989) *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Gries, S. T. and Stefanowitsch, A. (2007) *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis*. Berlin: Mouton de Gruyter.

Griffin, Z.M. (2004) The eyes are right when the mouth is wrong. *Psychological Science 15*(12): pp.814-821.

Griffin, Z.M. and Bock, K. (2000) What the eyes say about speaking. *Psychological science 11*(4): pp.274-279.

Gripsrud, J. (2002) *Understanding Media Culture*. London: Arnold.

Grivicic, T. and Nilep, C. (2004) When phonation matters: The use and function of yeah and creaky voice. *Colorado Research in Linguistics 17*(1): pp.1-11.

Grönqvist, L. (2004) Robust methods for automatic transcription and alignment of speech signals. *Speech and speaker recognition*, *Göteborg University* [online academic course material]. Available at: http://www.speech.kth.se/~matsb/speec h_rec_course_2003/ [Accessed 2006-02-10].

Gu, Y. (2006) Multimodal text analysis: A corpus linguistic approach to situated discourse. *Text and Talk 26*(2): pp.127-167.

Hadar, U. (1997) Interpreting at the surface . *Clinical Studies 3*: pp.83-104.

Hadar, U., Steiner, T.J. and Clifford Rose, F. (1985) Head movements during listening turns in conversation. *Behavioral Science 9*(4): pp.214-228.

Haiman, J. (1998) The metalinguistics of ordinary language. *Evolution of Communication 2*(1): pp.117-135.

Halliday, M.A.K. (1978) *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Arnold.

Halliday, M.A.K. (1992) A systemic interpretation of Peking syllable finals. In Tench, P. (Ed.) *Studies in systemic phonology*. London: Pinter. pp.98-121.

Halliday, M.A.K. and Hasan, R. (1976) *Cohesion in English.* London: Longman.

Heinz, B. (2003) Backchannel responses as strategic responses in bilingual speakers' conversations. *Journal of Pragmatics 35:* pp.1113-1142.

Hellerman, J. and Vergun, A. (2007) Language which is not taught: The discourse marker use of beginning adult learners of English. *Journal of Pragmatics 39*: pp.157-159.

Henley, N. and Kramarae, C. (1991) Miscommunication, Power, and Gender. In Coupland, N., Wiemann, J.M. and H. Giles (Eds.) *'Miscommunication' and Problematic Talk.* Newbury Park, CA: Sage. pp.18-43.

Heritage, J. (1984) A change-of-state token and aspects of its sequential placement. In Atkinson, J.M. and Heritage, J. (Eds.) *Structures of Social Interaction: Studies in Conversation Analysis.* Cambridge: Cambridge University Press. pp.299-345.

Heritage, J. (1998) Oh- prefaced responses to inquiry. *Language in Society 27*(3): pp.291-334.

Hill, D. R. (2000) Give us the tools: A personal view of multimodal computer–human dialogue. In Taylor, M.M., Ne´el, F. and Bouwhuis, G.G. (Eds.) *The Structure of Multimodal Dialogue II.* Amsterdam: John Benjamins. pp.25-62.

Hirschman, L. (1974) Analysis of supportive and assertive behaviour in conversations. *Proceedings of the Linguistic Society of America Conference.*

Holler, J. and Beattie, G. (2002) A micro-analytic investigation of how iconic gestures and speech represent core semantic features in talk. *Semiotica 142*(1-4): pp.31-69.

Holler, J. and Beattie, G. (2003) How iconic gestures and speech interact in the representation of meaning: Are both aspects really integral to the process? *Semiotica 146*(1-4): pp.81-116.

Holler, J. and Beattie, G.W. (2004) The interaction of iconic gesture and speech.
5th International Gesture Workshop, *Genova.* Heidelberg: Springer Verlag. pp.15-17.

Houtkoop, H. and Mazeland, H. (1985) Turns and discourse units in everyday conversation. *Journal of Pragmatics 9:* pp.595-619.

Huberty, C. J. (1993) Historical origins of statistical testing practices: The treatment of Fisher versus Neyman-Pearson views in textbooks. *The Journal of Experimental Education, 61*(4): pp.317-333.

Ide, N. (1998) Corpus encoding standard: SGML guidelines for encoding linguistic corpora. *First International Language Resources and Evaluation Conference.* Granada, Spain.

Ingram D. (1978) Sensori-motor development and language acquisition. In Lock, A. (Ed.) *Action, gesture and symbol: the emergence of language.* London: Academic Press.

Isard, M. and Blake, A. (1998) *CONDENSATION – conditional density propagation for visual tracking. International. Journal of Computer Vision 29*(1): pp.5-28.

Jefferson, G. (1984) Notes on a systematic deployment of the acknowledgement tokens "Yeah" and "Mm hm". *Papers in Linguistics 17*: pp.197-216.

Jefferson, G. (1993) Caveat speaker: preliminary notes on recipient topic-shift implicature. *Research on language and social interaction 26*: pp.1-30.

Jefferson, G. (2004) Glossary of transcript symbols with an introduction. In Lerner, G.H. (Ed.) *Conversation Analysis: Studies from the first generation*. Amsterdam/Philadelphia: John Benjamins. pp.13-31.

Käding J. (1879) *Häufigkeitswörterbuch der deutschen Sprache.* Steglitz: Privately Published.

Kanad, T., Cohn, J. and Tian, Y. (2000) Comprehensive database for facial expression analysis. *Proceedings of the International Conference of Face and Gesture Recognition* [online]. pp.46-53. Available at: http://ieeexplore.ieee.org/iel5/6770/18088/00840611.pdf [Accessed 22 October 2006].

Kapoor, A. and Picard, R.W. (2001) A Real-Time head nod and shake detector. *ACM International Conference Proceedings Series* [online]. pp.1-5. Available at: http://vismod.media.mit.edu/tech-reports/TR-544.pdf [Accessed 3 July 2006].

Kasper, G. (1995) Wessen Pragmatik? Für eine Neubestimmung sprachlicher Handlungskompetenz. *Zeitschrift für Fremdsprachenforschung 6:* pp.1-25.

Katz, J.S. and Martin, B.R. (1997) What is research collaboration? *Research Policy 26*: pp.1-18.

Kawato, S. and Ohya, J. (2000) Real-time detection of nodding and head shaking by directly detecting and tracking the 'between eyes'. *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, 2000* [online]. Available at: http://ieeexplore.ieee.org/iel5/6770/18088/00840610.pdf .ieee [Accessed 31 October 2006].

Kendon, A. (1967) Some functions of gaze-direction in social interaction. *Acta Psycholoigia 26*: pp.22-63.

Kendon, A. (1972) Some relationships between body motion and speech. In Seigman, A. and Pope, B. (Eds.) *Studies in Dyadic Communication*. Elmsford, New York: Pergamon Press. pp.177-216.

Kendon, A. (1979) Some emerging features of face-to-face interaction studies. *Sign Language Studies 22*: pp.7-22.

Kendon, A. (1980) Gesture and speech: Two aspects of the process of utterance. In Key, M.R. (Ed.) *Nonverbal Communication and Language*. New York: Mouton de Gruyter. pp.207-227.

Kendon, A. (1982) The organisation of behaviour in face-to-face interaction: observations on the development of a methodology. In Scherer, K.R. and Ekman, P. (Eds.) *Handbook of Methods in Nonverbal Behaviour Research*. Cambridge: Cambridge University Press. pp.440-505.

Kendon, A. (1983) Gesture and Speech: How they interact. In Wiemann, J. and Harrison, R. (Eds.) *Nonverbal Interaction*. California: Sage Publications. pp.13-46.

Kendon, A. (1987) On gesture: Its complementary relationship with speech. In Siegman, A.W. and Feldstein, S. (Eds.) *Nonverbal Behavior and Communication*. London: Lawrence Erlbaum Associates. pp.65-97.

Kendon, A. (1990) *Conducting Interaction.* Cambridge: Cambridge University Press.

Kendon, A. (1992) Some recent work from Italy on quotable gestures ('emblems'). *Journal of Linguistic Anthropology 2*(1): pp.77-93.

Kendon, A. (1994) Do gestures communicate? A review. *Research on Language and Social Interaction 27*(3): pp.175-200.

Kendon, A. (1997) Gesture. *Annual Review of Anthropology 26:* pp.109-128.

Kendon, A., Harris, R.M. and Key, M.R. (1976) *Organization of behavior in face-to-face interaction*. The Hague: Mouton.

Kennedy, G.D. (1998) *An Introduction to Corpus Linguistics.* London: Longman.

Kilgarriff, A. (1996) Which words are particularly characteristic of a text? A survey of statistical procedures. *AISB workshop of language engineering for document analysis and recognition, Sussex University*. pp.33-40.

Kilgarriff, A. (2005) Language is never, ever, ever random. *Corpus linguistics and Linguistic Theory 1*(2): pp.263-275.

Kipp, M. (2001) Anvil – A generic annotation tool for multimodal dialogue. *Proceedings of 7th European Conference on Speech Communication and Technology 2nd INTERSPEECH Event Aalborg, Denmark*. pp.1367–1370.

Kipp, M., Neff, M. and Albrecht, I. (2007) An annotation scheme for conversational gestures: how to economically capture timing and form. *Language Resources and Evaluation 41*(3/4): pp.325–339.

Kita, S., van Gijn, I. and van der Hulst, H. (1997) Movement Phase in Signs and Co-Speech Gestures, and Their Transcriptions by Human Coders. *Gesture Workshop, Germany* [online]. pp.23-35. Available at: http://www.informatik.uni-trier.de/~ley/db/conf/gw/gw1997.html [Accessed 16 November 2006].

Knight, D. (2006) Corpora: The Next Generation. Part of the AHRC funded online *Introduction to Corpus Investigative Techniques* [online]. Available at: http://www.corpus.bham.ac.uk/corpus-building.shtml [Accessed 1 December 2006].

Knight, D. and Adolphs, S. (2006) *Text, Talk and Corpus Analysis* [academic online module, restricted access]. University of Nottingham, England. pp.175-190.

Knight, D. and Adolphs, S. (2008) Multi-modal corpus pragmatics: the case of active listenership. In Romeo, J. (Ed.) *Corpus and Pragmatics*. Berlin and New York: Mouton de Gruyter.

Knight, D., Bayoumi, S., Mills, S., Crabtree, A., Adolphs, S., Pridmore, T. and Carter, R.A. (2006) Beyond the Text: Construction and Analysis of Multi-Modal Linguistic Corpora. *Proceedings of the 2nd International Conference on e-Social Science, Manchester, 28 - 30 June 2006* [online]. Available at: http://www.ncess.ac.uk/events/conference/2006/papers/abstr acts/KnightBeyondTheText.shtml [Accessed 14 September 2006].

Knight, D., Carter, R.A. and Adolphs, S. (2005) *The linguistic coding of backchannels: A proposed methodological approach* [unpublished NCeSS node meeting report]. University of Nottingham, England.

Knight, D., Evans, D., Carter, R.A. and Adolphs. S. (2009) Redrafting corpus development methodologies: Blueprints for 3rd generation multimodal, multimedia corpora. *Corpora.*

Knudsen, M.W., Martin, J.C., Dybkjær, L., Ayuso, M.J.M, Bernsen, N.O., Carletta, J., Kita, S., Heid, U., Llisterri, J., Pelachaud, C., Poggi, I., Reithinger, N., van ElsWijk, G. and Wittenburg, P. (2002) Survey of Multimodal Annotation Schemes and Best Practice. *ISLE Deliverable D9.1, 2002* [unpublished internal ISLE document].

Kopytko, R. (2003) What is wrong with modern accounts of context in linguistics?. *Vienna English Working Papers 12*: 45-60.

Krauss, R. M., Morrel-Samuels, P. and Colasante, C. (1991) Do conversational hand gestures communicate? *Journal of Personality and Social Psychology 61*: pp.743-754.

Krauss, R. M., Dushay, R. A., Chen, Y. and Rausher, F. (1995) The communicative value of conversational hand gestures. *Journal of Experimental Social Psychology 31*: pp.533-552.

Krauss, R.M. (2002) The psychology of verbal communication. In Smelser, N. and Baltes, P. (Eds.) *International Encyclopaedia of the Social and Behavioral Sciences*. London: Elsevier. pp.16161-16165.

Kress, G.R. and Van Leeuwen, T. (1996) *Reading Images*. London: Routledge.

Kress, G.R. and Van Leeuwen, T. (2001) *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London: Arnold.

Kruijff-Korbayová, V.R., Schehl, J., and Becker, T. (2006) The Sammie Multimodal Dialogue Corpus Meets the Nite XML Toolkit. *Proceedings of the Fifth Workshop on multi-dimensional Markup in Natural Language Processing (EACL) 2006* [online]. Available at: http://homepages.inf.ed.ac.uk/vrieser/papers/sammie-NLPXML06.pdf [Accessed 16 December 2008].

La Cascia, M., Sclaroff, S. and Athitsos, V. (2000) Fast, reliable head tracking under variable illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*(6): pp.322-326.

Labov, W. (1972) *Sociolinguistic Patterns.* Philadelphia, PA: University of Pennsylvania Press.

Labov, W. and Fanshel, D. (1977) *Therapeutic Discourse: Psychotherapy as Conversation*. New York: Academic Press.

Lapadat, J.C. and Lindsay, A. C. (1999) Transcription in research and practice: from standardisation of technique to interpretative positioning. *Qualitative Inquiry 5*(1): pp.64-86.

Laver, J. (1994) *Principles of phonetics*. Cambridge: Cambridge University Press.

Leech, G. (1991) The state of the art in corpus linguistics. In Aijmer, K. and Altenberg, B. (Eds.) *English Corpus Linguistics,* London: Longman. pp.8-39.

Leech, G. (1997) Introducing corpus annotation. In Garside, R., Leech, G. and McEnery, T. (Eds.) *Corpus annotation: Linguistic information from computer text corpora*. London: Longman. pp.1-18.

Leech, G. (2005) Adding Linguistic Annotation. In Wynne. M. (Ed.) *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbow Books. pp.1-16. Available online at: http://ahds.ac.uk/linguistic-corpora/ [Accessed 15 August 2006].

Leech, G., Myers, G. and Thomas, J. (Eds.) (1995) *Spoken English on Computer: Transcription, Mark-up and Application.* London: Longman.

Lenk, U. (1998) *Marking discourse coherence: Functions of discourse markers in spoken English*. Tübingen: Gunter Narr Verlag.

Lozano, S. C. and Tversky, B. (2006) Communicative gestures facilitate problem solving for both communicators and listeners. *Journal of Memory and Language 55*: pp.47-63.

Lund, K. (2007) The importance of gaze and gesture in interactive multimodal explanation. *Language Resources and Evaluation 41*(3): pp.289–303.

Malinowski, B. (1923) The Problem of Meaning in Primitive Languages. In Ogden, C.K. and Richards, I.A. (Eds.) *The Meaning of Meaning*. London: Routledge and Kegan Paul. pp.146-152.

Maltz, D. and Borker, R. (1982) A cultural approach to male-female miscommunication. In Gumperz, J. (Ed.) *Language and Social Identity*. Cambridge: Cambridge University Press. pp.196-216.

Mana, N., Lepri, B., Chippendale, P., Cappelletti, A., Pianesi, F., Svaizer, P., and Zancanaro, M. (2007) Multimodal Corpus of Multi-Party Meetings for Automatic Social Behavior Analysis and Personality Traits Detection. *Proceeding of Workshop on Tagging, Mining and Retrieval of Human-Related Activity Information at ICMI'07, Nagoya, Japan*. pp.9-14.

Markee, N. (2000) *Conversation Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Martinec, R. (1998) Cohesion in action. *Semiotica 120*(1/2): pp.161–180.

Martinec, R. (2001) Interpersonal resources in action. *Semiotica 135*(1/4): pp.117-145.

Mautner, G. (2007) Mining Large Corpora for Social Information: The Case of Elderly*. Language in Society 36*(1): 51–72.

Maynard, S.K. (1987) International functions of a nonverbal sign. Head movement in Japanese dyadic casual conversation. *Journal of Pragmatics 11*: pp.589-606.

Maynard, S.K. (1989) *Japanese Conversation: Self-contextualization through Structure and Interactional Management.* Norwood, New Jersey: Ablex.

Maynard, S.K. (1990) Conversation management in contrast: listener response in Japanese and American English. *Journal of Pragmatics 14*: pp.397-412.

Maynard, S.K. (1997) Analyzing interactional management in native/non-native English conversation: a case of listener response. *International Review of Applied Linguistics 35*(1): pp.37-60.

McCarthy, M.J. (2001) *Issues in Applied Linguistics*. Cambridge: Cambridge University Press.

McCarthy, M.J. (2002) Good listenership made plain: non-minimal response tokens in British and American spoken English. In Reppen, R., Fitzmaurice, S. and Biber, D. (Eds.) *Using corpora to explore linguistic variation*. Amsterdam: John Benjamins. pp.49-72.

McCarthy, M.J. (2003) Talking back: 'small' interactional response tokens in everyday conversation. *Research on Language and Social Interaction 36*(1): pp.33-63.

McClave, E.Z. (2000) Linguistic functions of head movements in the context of speech. *Journal of Pragmatics 32*(7): pp.855-878.

McCowan, S. Bengio, D. Gatica-Perez, G. Lathoud, F. Monay, D. Moore, P. Wellner, and Bourlard, H. (2003) Modelling Human Interaction in Meetings. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* Hong Kong, April 2003. pp.748-751.

McEnery, A. and Xiao, R. (2004) The Lancaster Corpus of Mandarin Chinese: A corpus for monolingual and contrastive language study. In Lino, M., Xavier, M., Ferreire, F., Costa, R. and Silva, R. (Eds.) *Proceedings of the*

Fourth International Conference on Language Resources and Evaluation (LREC) 2004. Lisbon: Portugal. pp.1175-1178.

McEnery, T. and Wilson, A. (1996) *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

McEnery, T., Xiao, R. and Tono, Y. (2006) *Corpus-based language studies- an advanced resource book*. London: Routledge.

McGregor, G. and White, R.S. (Eds.) (1990) *Reception and Response: Hearer Creativity and the Analysis of Spoken and Written Texts*. London: Routledge.

McNeill, D. (1979) *The Conceptual Basis of Language*. Hillsdale: Erlbaum.

McNeill, D. (1985) So you think gestures are nonverbal? *Psychological Review 92*(3): pp.350-371.

McNeill, D. (1992) *Hand and mind: What gestures reveal about thought*. Chicago: The University of Chicago Press.

McNeill, D., Cassell, J. and McCullough, K-E. (1994) Communicative effects of speech-mismatches gestures. *Research on Language and Social Interaction 27*(3): pp.223-237.

Meyer, C.F. (2002) *English corpus linguistics: An introduction.* Cambridge: Cambridge University Press.

Mishan, F. (2004) *Designing Authenticity into Language Learning Materials.* Bristol, GBR: Intellect Books.

Morimoto, C., Koons D., Amir A. and Flickner M. (1998) Real-time detection of eyes and faces. *Proceedings of the Workshop on Perceptual User Interfaces* [online]. pp.117-120. Available at: http://www.acm.org/icmi/1998/Papers/Morimoto.pdf [Accessed 14 December 2007].

Morris, C. (1946) *Signs, Language and Behaviour.* Englewood-Cliffs: Prentice Hall.

Morris, D., Collett, P., Marsh, P. and O'Shaughnessy, M. (1979) *Gestures*. New York: Stein and Day.

Mott, H. and Petrie, H. (1995) Workplace interactions: women's linguistic behaviour. *Journal of Language and Social Psychology 14*(3): pp.324-336.

Mulac, A., and Bradac, J. J. (1995) Women's style in problem solving interaction: Powerless or simply feminine? In Kalbfleisch, P.J. and Cody, M.J. (Eds.) *Gender, power, and communication in human relationships.* Hillsdale, NJ: Erlbaum. pp.83-104.

Mulac, A., Erlandson, K.T., Farrar, W.J., Hallett, J.S., Molloy, J.L. and Prescott, M.E. (1998) Uh-huh. What's that all about? Differing interpretation of conversational backchannels and questions as sources of miscommunication across gender boundaries. *Communication Research 25*(6): pp.641- 668.

Müller, C. (1996) *Gestik in Kommunikation und Interaction*. PhD thesis. Freie University: Berlin.

Myers, G.E. and Myers, T.T. (1973) *The Dynamics of Human Communication: A Laboratory Approach.* London: McGraw-Hill Book Company.

Newell, W.H. (1984) Interdisciplinary curriculum development in the 1970's: the paracollege at St. Olaf and the Western College Program at Miami University. In Jones, R.M. and Smith, B.L (Eds.) *Against the current: reform and experimentation in higher education*. Cambridge: Schenkman. pp.127–47.

Newton, E.M., Sweeney, L. and Malin, B. (2005) Preserving Privacy by De-Identifying Face Images. *IEEE Transactions on Knowledge and Data Engineering 17*(2): pp.232-243.

Nobe, S. (1996) *Cognitive rhythms gestures and acoustic aspects of speech.* PhD thesis. Department of Psychology: University of Chicago, Illinois.

Noller, P. (1984) *Nonverbal Communication and Marital Interaction.* Oxford: Pergamon Press.

Norris, S. (2004) *Analysing Multimodal Interaction: A Methodological Framework.* London: Routledge.

O'Connell, D.C. and Kowal, S. (1999) Transcription and the issue of standardisation. *Journal of Psycholinguistic research 28*(2): pp.103-120.

Ochs, E. (1979) Transcription as theory. In Ochs, E. and Schieffelin, B.B. (Eds.) *Developmental Pragmatics.* New York: Academic Press. pp.43-72.

O'Keeffe, A. and Adolphs, S. (2008) Using a corpus to look at variational pragmatics: Response tokens in British and Irish discourse. In Schneider, K.P. and Barron, A. (Eds.) *Variational Pragmatics.* Amsterdam, Netherlands: John Benjamins. pp.69-98.

O'Keeffe, A., McCarthy, M. and Carter, R. (2007) *From corpus to classroom-Language use and language teaching.* Cambridge: Cambridge University Press.

Ong, S. and Ranganath, S. (2005) Automatic sign language analysis: a survey and the future beyond lexical meaning. *IEEE Transactions on Pattern Analysis and Machine Intelligence 27*(6): pp.873-891.

Oreström, B. (1983) *Turn-taking in English Conversation.* Lund, Sweden: LiberFörlag Ltd.

Palmer, H. (1933) *Second interim report on English collocations.* Tokyo: Institute for Research in English Teaching.

Pantic, M. and Rothkrantz, L. (1999) An expert system for multiple emotional classification of facial expressions. *Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence* [online]. pp.113-120. Available at: http://ieeexplore.ieee.org/xpl/RecentCon.jsp?punumber=6582 [Accessed 29 August 2007].

Pantic, M. and Rothkrantz, L. (2000) Expert system for automatic analysis of facial expression. *Image and Vision Computing 18*(11): pp.881-905.

Patton, B.R. and Giffin, K. (1981) *Interpersonal Communication in Action-Basic Texts and Readings.* Cambridge: Harper and Row.

Pea, R., Mills, M., Rosen, J., Dauber, K., Effelsberg, W. and Hoffert, E. (2004) The diver project: Interactive digital video repurposing. *IEEE MultiMedia 11*(1): pp.54–61.

Pedhazur, E.J., and Schmelkin, L.P. (1991) *Measurement, design, and analysis: An integrated approach.* Hillsdale, NJ: Erlbaum.

Pittner, S. and Kamarth, S. (1999) Feature extraction from wavelet coefficients for pattern recognition tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*(1): pp.83-88.

Prillwitz, S., Leven, R., Zienert, H., Hanke, T. and Henning, J. (1989) *HamNoSys. Version 2.0. Hamburg Notation System for Sign Language. An Introductory Guide.* Hamburg: Signum.

Psathas, G and Anderson, T. (1990) The 'practices' of transcription in conversation analysis. *Semiotica 78*(1/2): pp.75-99.

Rayson, P. (2003) *Matrix: A statistical method and software tool for linguistic analysis through corpus comparison*. PhD thesis [online]. Lancaster University. Available at: http://eprints.comp.lancs.ac.uk/753/1/phd2003.pdf [Accessed 1 February 2006].

Reppen, R. and Simpson, R. (2002) Corpus linguistics. In Schmitt, N. (Ed.) *An Introduction to Applied Linguistics*. London: Arnold. pp.92-111.

Richmond, V.P., McCroskey, J.C. and Payne, S.K. (1991) *Nonverbal Behaviour in Interpersonal Relations*. Prentice Hall: New Jersey.

Rimé, B. (1982) The elimination of visible behaviour from social interactions: effects on verbal, nonverbal and interpersonal variables. *European Journal of Social Psychology 12*: pp.113-129.

Rimé, B. and Schiarartura, L. (1991) Gesture and speech. In Feldman, R.S. and Rimé, D. (Eds.) *Fundamentals of Nonverbal Behaviour*. New York: Cambridge University Press. pp.239-284.

Roberts, C. (2006) Part one: issues in transcribing spoken discourse. *Qualitative research methods and transcription* [online academic course]. Kings College London. Available at: http://www.kcl.ac.uk/schools/sspp/ education/research/projects/dataqual.html [Accessed 10 March 2007].

Roger, D. and Nesshover, W. (1987) Individual differences in dyadic conversation strategies: a further study. *British journal of Social Psychology 26*: pp.247-255.

Roger, D., Bull, P. and Smyth, S. (1988) The development of a comprehensive system for classifying interruptions. *Journal of Language and Social Psychology 7*: pp.27-34.

Roger, D.B. and Schumacher, A. (1983) Effects of individual differences on dyadic conversational strategies. *Journal of Personality and Social Psychology 45*: pp.700-705.

Rosenberg, E.L., Ekman, P., Jiang, W., Coleman, R.E., Hanson, M., O'Connor, C., Waugh, R., and Blumenthal, J.A. (2001) Linkages between facial expressions of anger and transient myocardial ischemia in men with coronary artery disease. *Emotion 1*: pp.107-115.

Rost, M. (2002) *Teaching and Researching Listening*. London, UK: Longman.

Sacks, H., Schegloff, E.A. and Jefferson, G. (1974) A simplest systematics for the organisation of turn taking for conversation. *Language 50*(4-1): pp.696-735.

Saferstein, B. (2004) Digital technology- methodological adoption: Text and video as a resource for analytical reflectivity. *Journal of Applied Linguistics 1*(2): pp.197-223.

Schegloff, E. (1972) Notes on a conversational practice: formulating place. In Sudnow, D.N. (Ed.) *Studies in social interaction*. New York: Free Press.

Schegloff, E. (1982) Discourse as interactional achievement: some uses of "uh huh" and other things that come between sentences. In Tannen, D. (Ed.) *Analyzing Discourse, Text, and Talk.* Washington DC: Georgetown University Press. pp.71-93.

Schegloff, E.A. (1984) On some gestures' relation to talk. In Atkinson, J.M. and Heritage, E.J. (Eds.) *Structures of Social Action: Studies in Conversation Analysis.* Cambridge: Cambridge University Press. pp.266-296.

Schiel, F. and Mögele, H. (2008) Talking and Looking: the SmartWeb Multimodal Interaction Corpus. *Proceedings of Sixth International*

Conference on Language Resources and Evaluation (LREC) 2008 [online]. Available at: http://www.lrec-conf.org/proceedings/lrec2008/ pdf/510_paper.pdf [Accessed 16 December 2008].

Schiel, F., Steininger, S. and Türk, U. (2002) The SmartKom Multimodal Corpus at BAS. *Proceedings of the 3rd Language Resources and Evaluation Conference (LREC) 2002* [online]. Available at: https://www.phonetik.uni-muenchen.de/ forschung/publikationen/Schiel-02-SKCorpus.ps [Accessed 16 December 2008].

Schiffrin, D. (1994) *Approaches to discourse.* Oxford: Blackwell.

Scholfield, P. (1995) *Quantifying Language.* Clevedon: Multilingual Matters Ltd.

Scollon, R. (1998) *Mediated Discourse as Social Interaction.* London: Longman.

Scott, M. (1999) *Wordsmith Tools* [Computer program]. Oxford: Oxford University Press.

Shannon, C.E. and Weaver, W. (1949) *A Mathematical Model of Communication.* Urbana, IL: University of Illinois Press.

Sidner, C.L., Lee, C., Morency, L-P. and Forlines, C. (2006) The Effect of Head-Nod Recognition in Human-Robot Conversation. *Proceedings of the ACM SIGCHI/SIGART Conference on Human-Robot Interaction (HRI)* [online]. pp.290-296. Available at: http://people.ict.usc.edu/~morency/ Papers/hri06.pdf [Accessed 4 December 2007].

Sinclair, J. (1996) The search for units of meaning. *Textus 9*(1)*:* pp.71-106.

Sinclair, J. (2004) *Trust the text: Language, Corpus and Discourse.* London: Routledge.

Sinclair, J. (2005) Corpus and text- basic principles. In Wynne, M. (Ed.) *Developing Linguistic Corpora: a Guide to Good Practice.* Oxford: Oxbow Books. pp.1-16. Available online at: http://ahds.ac.uk/linguistic-corpora/ [Accessed 2 October 2006].

Sinclair, J. (2008) Borrowed ideas. In Gerbig, A. and Mason, O. (Eds.) *Language, people, numbers- Corpus Linguistics and society.* Amsterdam: Rodopi BV. pp.21-42.

Sperberg-McQueen, C.M. and Burnard, L. (1999) *Guidelines for electronic text encoding and interchange (TEI P3)* Chicago and Oxford: ACH-ALLC-ACL Text Encoding Initiative.

Stenström, A.B. (1987) Carry on symbols in English conversation. In Meijs, W. (Ed.) *Corpus Linguistics and Beyond: Proceedings of the Seventh International Conference on English Language Research on Computerized Corpora.* Amsterdam: Rodopi. pp.87-119.

Strassel, S. and Cole, A.W. (2006) Corpus development and publication. *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC) 2006* [online]. Available at: http://papers.ldc.upenn.edu/LREC 2006/CorpusDevelopmentAndPublication.pdf [Accessed 4 April 2007].

Streeck, J. (1993) Gesture as Communication 1: Its coordination with gaze and speech. *Communication Monographs 60*(4): pp.275-299.

Streeck, J. (1994) Gestures as Communication 2: The audience as co-author. *Research on Language and Social Interaction 27*(3): pp.239-267.

Stubbe, M. (1998a) Are you listening? Cultural influences on the use of supportive verbal feedback in conversation. *Journal of Pragmatics 29*: pp.257-289.

Stubbe, M. (1998b) Researching language in the workplace: A participatory model. *Proceedings of the Australian Linguistics Society Conference* [online]. Available at: http://emsah.uq.edu.au/linguistics/als/als98/ [Accessed 10 May 2007].

Stubbs, M. (1994) Grammar, text an ideology: computer-assisted methods in the linguistics of representation. *Applied Linguistics 15*(2): pp.201-223.

Stubbs, M. (1996) *Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture*. Oxford: Blackwell.

Tao, H. and Thompson, S. (1991) English backchannels in Mandarin conversations: A case study of superstratum pragmatic 'interference'. *Journal of Pragmatics 16*. pp.209–233.

Taylor, M.M., Ne´el, F. and Bouwhuis, G.G. (Eds.) (1999) *The Structure of Multimodal Dialogue II*. Amsterdam: John Benjamins.

ten Have, P. (2007) *Doing Conversational Analysis: A practical guide*. 2nd ed. London: Sage.

Thomas, J. (1983) Cross-cultural pragmatic failure. *Applied Linguistics 4*(2): pp.91-122.

Thompson, L.A. and Massaro, D.W. (1986) Evaluation and integration of speech and pointing gestures during referential understanding. *Journal of Experimental Child Psychology 42*(1): pp.144-168.

Thompson, P. (2005) Spoken Language Corpora. In Wynne, M. (Ed.) *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbow Books. pp.59-70. Available online at: http://ahds.ac.uk/linguistic-corpora/ [Accessed 2 October 2006].

Thorne, B. and Henley, N. (Eds.) (1975) *Language and Sex: Difference and Dominance*. Rowley, MA: Newbury House.

Thoutenhoofd, E.D. (2007) Corpus linguistics as multimedia laboratory: Material culture and experimental practice in the social sciences. Paper presented at the *Science and technology studies views of e-social science panel, Third International Conference on e-Social Science*, Ann Arbor: USA.

Tian, Y., Kanade, T. and Cohn, J.F. (2000) Dual-state parametric eye tracking. *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, March 2000* [online]. pp.110-115. [Accessed 1 June 2006].

Tottie, G. (1991) Conversational style in British and American English: The case of backchannels. In Aijmer, K. and Altenberg, B. (Eds.) *English corpus linguistics*. London: Longman. pp.254-271.

Trojanová, J., Hrúz, M., Campr, P. and Železný, M. (2008) Design and Recording of Czech Audio-Visual Database with Impaired Conditions for Continuous Speech Recognition. *Proceedings of Sixth International Conference on Language Resources and Evaluation (LREC) 2008* [online]. Available at: http://www.lrec-conf.org/proceedings/lrec2008/pdf/316_paper.pdf [Accessed 16 December 2008].

Van Dijk, T.A. (Ed.) (1977) *Text and context: explorations in the semantics and pragmatics of discourse*. London; Longman.

van Son, R.J.J.H., Wesseling, W., Sanders, E., and van der Heuvel, H. (2008) The IFADV corpus: A free dialog video corpus. *Proceedings of Sixth International Conference on Language Resources and Evaluation (LREC) 2008* [online]. Available at: http://www.lrec-conf.org/proceedings/lrec2008/pdf/132_ paper.pdf [Accessed 16 December 2008].

Vermeerbergen, M. (2006) Past and current trends in sign language research. *Language and Communication 26*(2): pp.168-192.

Watzlawick, P., Beavin, J. and Jackson, D. (1967) *Pragmatics of Human Communication*. W. W. Norton: New York.

White, S. (1989) Backchannels across cultures: A study of Americans and Japanese. *Language in Society 18*: pp.59-76.

Widdowson, H.G. (2000) On the Limitations of Linguistics Applied. *Applied Linguistics 21*(1): 3–25.

Wilcox, S. (2004) Language from gesture. *Behavioral and Brain Sciences 27*(4): pp.524-525.

Wittenburg, P., Broeder, D. and Sloman, B. (2000) Meta-description for language resources. *EAGLES/ ISLE White paper* [online report]. Available at: http://www.mpi.nl/world/ISLE/documents/papers/white_paper_11.pdf. [Accessed 2 October 2006].

Wolf, J.C. and Bugmann, G. (2006) Linking Speech and Gesture in Multimodal Instruction Systems. *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication* (RO-MAN06) 2006 [online]. pp.141-144. Available at: http://www.tech.plym.ac.uk/soc/staff/guidbugm/pub/ro-man-06_wolf_bugmann.pdf [Accessed 10 November 2008].

Wray, A., Trott, K. and Bloomer, A. (1998) *Projects in Linguistics: A practical guide to researching language.* London: Arnold.

Wynne, M. (2005) Archiving, distribution and preservation. In Wynne, M. (Ed.) *Developing Linguistic Corpora: a Guide to Good Practice.* Oxford: Oxbow Books. pp.71-78. Available online at http://ahds.ac.uk/linguistic-corpora/ [Accessed 2 October 2006].

Yngve, V.H. (1970) On getting a word in edgewise. *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society:* pp.567-577.

Železný, M., Krňoul, Z., Císař, P. and Matoušek., J. (2006) Design, implementation and evaluation of the Czech realistic audio-visual speech synthesis. *Signal Processing 83*(12): pp.3657-3673.

Zipf, G. K. (1935) *The Psychobiology of Language*. Boston: Houghton Mifflin.