

**STUDIES OF THE
PHONETICS-INTERACTION INTERFACE:
CLICKS AND INTERACTIONAL STRUCTURES
IN ENGLISH CONVERSATION**

MELISSA WRIGHT

PhD

2005

ABSTRACT

This thesis investigates the interface between the phonetic and the interactional organisation of naturally-occurring British and American English conversation. It demonstrates that clicks—previously regarded as having only a paralinguistic function in English—do in fact have an orderly, sequential distribution which can be mapped onto the interactional structure of talk. Moreover, clicks are shown to have different functions according to their embedded contexts of production.

A novel methodological approach is employed which combines (1) the sequential techniques of Conversation Analysis (CA), and (2) the phonetic techniques of impressionistic observation and instrumental analysis. CA enables the implementation of a Firthian approach, which regards language as being polysystemic, and its meaning as being inseparable from its context. A key feature of the methodology adopted in this thesis is the study of naturally-occurring conversation, rather than intuited or laboratory speech data.

Three separate analyses are presented that concern the phonetic-linguistic and sequential-interactional properties of (1) Multi-Unit, First-Closing Turns - which speakers employ to move into the closing section of a telephone call (2) New Sequence Indexing Clicks - which function to demarcate disjunctive sequences of talk and (3) Word Search Clicks - which undertake a turn-holding function. In all these analyses, systematic mappings are evidenced between the phonetic, sequential and interactional designs.

The findings of this research make significant contributions to our understanding of the linguistic-phonetic and interactional organisation of everyday social intercourse. They demonstrate the fruitfulness of implementing context-bound phonetic investigations alongside interactional analyses, since previously unobserved orderly and recurrent patterns in the phonetics-interaction interface can be clearly identified.

LIST OF CONTENTS

Abstract	ii
List of Contents	iii
List of Figures	vii
List of Tables	x
Acknowledgments	xi
Author's Declaration	xii

INTRODUCTION	1
---------------------------	----------

CHAPTER 1: CONTEXT

1.1 Introduction	3
1.2 The Firthian Approach	4
1.2.1 Context of Situation	4
1.2.2 Multi-Level Statements of Meaning	7
1.2.3 Rejection of Psychological Phenomena	9
1.2.4 Summary	10
1.3 Variationist Sociolinguistic Approach	12
1.3.1 Macro Contextual Feature: Social Attributes of Speaker	13
1.3.2 The Linguistic Variable	16
1.3.3 Quantification	17
1.3.4 Summary	19
1.4 Conversation Analysis (CA): A Firthian Toolbox	21
1.4.1 Talk-in-Interaction as Social Action	23
1.4.2 Actions are Context-Shaped and Context-Renewing	23
1.4.3 Polysystemicity	25
1.4.4 Rejection of Intuited and Experimental Speech Data	27
1.4.5 Participant Orientation as a Basis for Theorising	28
1.4.6 Summary	30
1.5 Invocations of Context in CA and Variationist Studies	31
1.5.1 Distal vs. Proximate	31
1.5.2 Evidencing Proximate and Distal Contextual Features	34
1.5.3 Summary	40
1.6 Directions of and Motivations for Research	40
1.7 Outline of Thesis	48

CHAPTER 2: DATA AND METHODOLOGY

2.1 Introduction.....	51
2.2 Data.....	51
2.2.1 The Corpora.....	51
2.2.2 Transcription Conventions.....	52
2.3 Conversation Analysis.....	54
2.3.1 Participant Orientations as Evidence.....	56
2.3.2 Participant Orientations as a Measure of Significance.....	59
2.4 Phonetic Approach.....	62
2.4.1 Impressionistic Analysis.....	62
2.4.2 Instrumental Analysis.....	64
2.5 Summary.....	68

CHAPTER 3: MULTI-UNIT, FIRST-CLOSING TURNS

3.1 Introduction.....	70
3.1.1 Previous Research on Call Closings.....	72
3.1.2 Outline of Study.....	75
3.2 Interactional Properties of Multi-Unit, First-Closing (F1) Turns.....	77
3.2.1 The Multi-Unit, Multi-Action Status of F1 turns.....	78
3.2.2 Recipient Provides Acceptance Marker.....	81
3.2.3 Caller Typically Produces F1 Turn.....	87
3.2.4 Summary.....	88
3.3 Lexical Properties of Multi-Unit, First-Closing Turns.....	88
3.3.1 First Unit Typically Comprised of One Lexical Item.....	89
3.3.2 ‘Okay’ Routinely Situated as First Lexical Item in Second Unit and Typically Followed by ‘then’ as in ‘okay then’.....	93
3.3.3 Summary.....	95
3.4 Phonetic Properties of Multi-Unit, First-Closing Turns.....	96
3.4.1 Differences in Placement of Units Relative to Speakers’ Pitch Range.....	97
3.4.2 Differences in Relative Pitch Spans of Units.....	98
3.4.3 Differences in Unit-Final Pitch Contours.....	101
3.4.4 Complete Closure at End of Final Syllable of Initial Unit.....	102
3.4.5 Pitch Upstep in Onset of Second Unit.....	107
3.4.6 Glottalisation in Onset of Second Unit.....	110
3.4.7 Differences in Amplitude between Units.....	111
3.4.8 Inter-Unit Clicks.....	113
3.5 Summary and Discussion.....	116

CHAPTER 4: CLICKS AND SEQUENCE MANAGEMENT

4.1 Introduction.....	119
4.1.1 Outline of Study.....	121
4.2 Phonetic Properties of Clicks in the Collection.....	122
4.3 Sequential Environment of New Sequence Indexing (NSI) Clicks.....	127
4.3.1 Minimal Adjacency Pairs in Opening Sequences.....	128
4.3.2 Assessments.....	131
4.3.3 Figurative Expressions.....	135
4.3.4 Sequence Closing Repetitions.....	138
4.3.5 Minimal Final Closing Tokens.....	142
4.3.6 Prefatory Discontinuity Markers.....	144
4.3.7 Summary.....	146
4.4 Participant Orientations to NSI Clicks Turns.....	146
4.4.1 Click-Initiated New Sequences Accepted by their Recipients.....	147
4.4.2 Click Recipients can Produce Subsequent New Sequence.....	149
4.4.3 Simultaneous Move to New Sequence by Both Participants.....	152
4.4.4 Summary.....	154
4.5 Phonetic Properties of NSI Click Turns.....	154
4.5.1 Differences in Pitch Placement of Pre and Post Click Sequences.....	155
4.5.2 Differences in Relative Pitch Spans of Pre and Post Click Sequences.....	156
4.5.3 Closure Maintained between Pre Click Sequence and Click in Multi Unit NSI turns.....	159
4.5.4 Upstep in Pitch in Onset of Click-Initiated New Sequences.....	162
4.5.5 Glottalisation in Onset of Click-Initiated New Sequences.....	165
4.5.6 Increase in Loudness in Onset of Click-Initiated New Sequences.....	167
4.5.7 Loud Inbreaths in Click-Initiated New Sequences.....	169
4.5.8 Particle ‘Uhm’ in Click-Initiated New Sequences.....	169
4.6 Summary and Discussion.....	173

CHAPTER 5: CLICKS AND WORD SEARCHES

5.1 Introduction.....	176
5.1.1 Word Searches in Talk-in-Interaction.....	177
5.1.2 Previous Research on Word Searches.....	180
5.1.3 Difficulties in Identifying Word Searches.....	181
5.1.4 Outline of Study.....	184
5.2 Pre Word Search Stretch.....	188
5.2.1 Incompleteness of Pre Word Search Stretch.....	188
5.2.2 Pre Word Search Projections of Searched-For Item.....	193
5.2.3 Summary.....	195
5.3 Word Search Stretch.....	196
5.3.1 Production of Clicks in Word Search Stretch.....	196
5.3.2 Particle ‘Uhm’ Initiates Word Search Stretch.....	198

5.3.3 Wh-Questions as Overt Word Search Displays.....	200
5.3.4 Occurrence of Clicks before Wh-Question or Searched-For Item	208
5.3.5 Glottal Closure Typically Maintained between Click and Onset of Searched-For Item.....	210
5.3.6 Particle ‘Uhm’ always situated before Click.....	212
5.3.7 Closure Typically Maintained between ‘Uhm’ and Click.....	213
5.3.8 Summary.....	215
5.4 Post Word Search Stretch	216
5.4.1 Completion of Turn Construction Unit	216
5.4.2 Pitch Match between Offset of Pre Word Search Stretch and Searched-For Item	218
5.4.3 Summary.....	219
5.5 Recipient Orientations to Word Searches	220
5.6 Summary and Discussion.....	225

CHAPTER 6: SUMMARY AND CONCLUSIONS

6.1 Introduction.....	229
6.2 Summary of Thesis and Findings.....	229
6.2.1 Summary of Chapters	229
6.2.2 Summary of Main Findings.....	232
6.2.3 Summary of Findings on Clicks	233
6.3 Limitation of Study and Directions for Further Research.....	238
6.4 Implications of Study	240
6.4.1 Clicks and Paralinguistics.....	241
6.4.2 Polysystemic Approaches to Language are Needed.....	241
6.4.3 Context-Embedded, Multi-Level Investigations of Talk are Profitable ...	242
6.4.4 Both Form-First and Function-First Analyses are Fruitful.....	244
6.4.5 Applications.....	245
6.5 Final Remarks	246

APPENDICES

Appendix 1: Transcription Conventions	248
Appendix 2: Details of Corpora	249

References.....	251
-----------------	-----

LIST OF FIGURES

1.1	Illustration of Variationist view of influences on speech production	12
3.1	Pitch trace to show relative placements of first and second units of F1 turns	98
3.2	Pitch trace to show relative spans of first and second units of F1 turns	99
3.3	Graph to show pitch placements and spans of first and second units of all F1 turns	100
3.4	Pitch trace to show shape of unit-final pitch contour in first and second units of F1 turns	101
3.5	Wave form and spectrogram to show closure at end of first unit in F1 turn	104
3.6	Wave form and spectrogram to show friction and bilabial closure in coda of first unit of F1 turn	106
3.7	Pitch trace to show upstep in onset of second unit of F1 turn relative to offset of first	109
3.8	Graph to show upstep in onset of second unit relative to offset of first in all F1 turns	109
3.9	Wave form and intensity trace to show upstep in intensity of onset of second unit of F1 turn relative to offset of first unit	112
3.10	Expanded wave form and spectrogram to show inter-unit bilabial click	115
4.1	Wave form and spectrogram to show affricated release of voiceless velar closure	123
4.2	Wave form and spectrogram to show non-affricated release of voiceless velar closure	123
4.3	Intensity trace, wave form and spectrogram to show high amplitude of click relative to following talk	125
4.4	Schema of typical sequence composition of call openings which precede NSI click turns	129
4.5	Schema of sequence-closing assessments which precede NSI click turns	132
4.6	Schema of sequence-closing figurative expressions which precede NSI click turns	136
4.7	Schema of sequence-closing repetitions which precede NSI click turns	139
4.8	Schema of sequence in which click recipient produces a new sequence in	149

	overlap with click-producers' subsequent inbreath	
4.9	Schema of sequence in which click recipient produces a new sequence after a minimal final closing token and following gap	150
4.10	Pitch trace to show placements of pre and post click sequences	156
4.11	Pitch trace to show relative spans of pre and post click sequences	157
4.12	Graph to show relative pitch placements and spans of first and second units of multi-unit NSI click turns	157
4.13	Pitch trace to show low placement of sequence-closing figurative expression	158
4.14	Pitch trace to show relative differences between placement of pre and post click talk	159
4.15	Wave form and spectrogram to show portion of 'silence' between offset of pre-click and following click-initiated new sequence	162
4.16	Pitch trace to show upstep in onset of click-initiated new sequence relative to offset of pre-click sequence	163
4.17	Pitch trace to show upstep in onset of click-initiated new sequence relative to offset of pre-click sequence	163
4.18	Graph to show upstep in pitch in onset of click-initiated new sequence relative to offset of preceding sequence	164
4.19	Pitch trace to show upstep in onset of post click sequence relative to offset of pre-click sequence in environment where non-click producer initiates post click shift in sequence	165
4.20	Wave form and spectrogram to show glottalisation in onset of first lexical item of click-initiated new sequence	166
4.21	Wave form and intensity trace to show upstep in intensity of onset of click-initiated new sequence relative to preceding sequence	168
4.22	Wave form and spectrogram to show high amplitude inbreath after click	169
4.23	Pitch trace to show pitch design of particle 'uhm'	171
4.24	Wave form and spectrogram to show initial glottality in onset of particle 'uhm'	172
5.1	Illustration of the different components of word search in fragment 2	184
5.2	Illustration of the different components of word search in fragment 4	185
5.3	Pitch trace to show incomplete prosody at 'uhm' and focal accent on 'Brendan'	191
5.4	Wave form and spectrogram to show clicks produced in fragment 4	198
5.5	Wave form and pitch trace to show parenthetical design of the oh-prefaced wh-question in fragment 14	206
5.6	Wave form and pitch trace to show parenthetical design of the oh-	207

	prefaced wh-question in fragment 4	
5.7	Wave form and spectrogram to show immediate onset of oh-prefaced wh-question after click in fragment 14	209
5.8	Wave form and spectrogram to show bilabial click produced before production of searched-for item in fragment 11	210
5.9	Pitch trace, wave form and intensity trace to show incomplete prosody at 'uhm' and focal accent on searched-for item 'Brendan'	218
6.1	Representation of new sequence indexing clicks and their sequential environment	234
6.2	Representation of word search clicks and their sequential environment	235
6.3	Representation of the disjunctive sequence junction piece and the conjunctive word search piece	236

LIST OF TABLES

3.1	Semitone spans of first and second units in F1 turns in fragments 1-9	99
3.2	Number of semitones in upstep in pitch in onset of the second unit in fragments 1-9	108
3.3	Increase in loudness in onset of second unit of F1 turn relative to first	111
4.1	NSI click turns which comprise multi-units	142
4.2	Prefatory discontinuity markers in NSI click turns	145
4.3	Recipients' fitted responses to NSI click turns produced in fragments 1-20	147
4.4	Number of decibels in upstep of click-initiated new sequence relative to offset of preceding sequence	167

ACKNOWLEDGEMENTS

To provide thanks to all the people who have helped me throughout the many varied stages of this research would require the writing of another thesis. As I plan to regain life in all its glory now that my thesis is done, I won't undertake the writing of another just yet (Crispin will be pleased to know that!). In what follows, then, I hope to provide at least adequate thanks to the people who have played a key role in helping me get to this wonderful final stage of saying thanks. To those not mentioned, thanks to you too.

Within the academic arena, I would firstly like to say a big thank you to Richard Ogden for supervising this thesis and for providing me with much encouragement, guidance and laughter throughout. Thanks also to my thesis advisory panel – John Local and Ros Temple: I thank John for his inspiration and help in organising my ideas, and Ros for her encouragement and for believing in me. Warm thanks are also due to Elizabeth Couper-Kuhlen and John Local for providing me with such an interesting, stimulating and enjoyable viva. I also thank the Arts and Humanities Research Board for funding this study.

Outside of academia I would like to thank my family for their wonderful love, support and patience, and for putting up with me being 'chained' to my PhD for so long. Thanks to my Dad and (step) Mum, without whose generosity, postgraduate research would probably never have been possible. Thanks to them also for providing me with support, sunshine and smiles. Dad will be shocked to know that this study deals mainly with 'one letter' rather than the previous 'three' I dealt with in my MA: a new joke for the Christmas newsletter perhaps?! Massive thanks also go to my Mum and Robbie who have also been never ending in their kindness and support, and who have stuck by me even when they weren't sure what I was doing or why I was doing it. They too have provided many smiles and comforts and have both been wonderful in every way. Special thanks to Mum for all the natters, even though we may have given each other 'cauliflower ear' on more than one occasion! Thanks also to Abbie and (my fab brother in law!) Diarmuid, who have always provided perfect escapes and entertaining breaks from the research. Special thanks to Abbie for being both my perfect big sis and friend. I still wonder how many of her mates think I study phrenology though! Thanks also to Rachel (for all the humorous stories!) and Darren, and to Mike and Sam. No family thanks would be complete without mentioning my fabulous Auntie Brenda, and to her I am thankful that she is my Auntie, and that she is who she is.

Moving onto my friends, I should also start by saying thanks for putting up with me with not being around for so long and for not ever taking my absence personally. Special thanks go to: Karen for the perspective on life, Anna and Dom for the silliness - with extra thanks to Anna for her friendship and understanding; Nick for his energy and refreshingness, Beth for hanging in there, Kristín for the boogies, Adam for the Mahjong, Rebekka for the dreams, Dave P for ‘happy Daves’, Gayan for the encouragement, the Wentworth gang for the good times, and Jennie, Rosa and Anna ‘the loud trombonist’ for the giggles. Big thanks also to the ‘motley crew’—Lola, Dave, Carrie, Vito, Kat and Kash—who have each and collectively provided me with some fantastic times and who have often unshackled me from my thesis. Extra thanks go to Lola for being such a great friend throughout. Thanks also to my wonderful next door neighbour and friend, Dorothy, who has provided me with lots of humour and welcome breaks in the garden. Warm thanks also to Laurian, Michael and Philippa for their kindness and support.

Thanks are also due to everyone in the department—staff and students—who have helped make my time in York so interesting and fun. Life in the department would not have been the same were it not for the office staff: thanks in particular to Judy for always patiently helping me with my faff and special thanks to Cynthia—my surrogate Mum!—whose chats always provided a nice change from the study. My office mates were also a great support: thanks in particular to Hazel for the biscuits and tiffin breaks, and to Helen for her loveliness. Thanks also to Plantie and Fishies for always looking calm and gorgeous!

Finally, I want to thank Crispin, who has held my hand through all the ups, downs and spirals of this research and who has given an endless (and very patient) listening ear to all my PhD ramblings. Thanks to him for his enthusiasm, love and laughter.

AUTHOR’S DECLARATION

I hereby declare that all material presented in this thesis is my own work unless otherwise stated. Some parts of chapter 3 have previously been presented (see Wright 2004).

INTRODUCTION

It is striking, and perhaps alarming, that even though the most predominant use of spoken language is that of everyday, naturally-occurring social interaction, knowledge and understanding of the linguistic-phonetic organisation of speech housed in its natural habitat of talk-in-interaction remains relatively limited. Moreover, current awareness and comprehension of the interactional implications and significances of certain phonetic events in naturally-occurring talk-in-interaction remains even more limited. These continued limitations in knowledge and comprehension can, in part, be attributed to the overemphasis in much linguistic-phonetic research of the analysis of contrived, decontextualised or intuited speech data. Such data are routinely employed in an attempt to regularise some of the variability and ‘messiness’ regarded as typifying contextualised, naturally-occurring speech productions. Where naturally-occurring, contextualised talk is examined, there is a tendency for the categories of analysis to be intuited rather than generated from and grounded within the data itself; thus these categories and their phonetic exponents typically bear little relevance when reapplied to conversational data.

Recent years have seen a growing number of linguists address these limitations in knowledge and understanding by combining (a) the interactional analysis techniques of Conversation Analysis (CA) and (b) the phonetic techniques of impressionistic and instrumental analysis to explore the phonetics-interaction interface of naturally-occurring talk-in-interaction (see e.g. Local & Kelly 1986; Local et al. 1986; Local 1992, 1996, 2004; Couper-Kuhlen 1993, 1996, 2004a, 2004b; Tarplee 1993; Wells & Peppé 1996; Selting 1996, 2000; Wells & Macfarlane 1998; Ogden 2001, 2003, 2004; Wright 2001, 2002a, 2004; Fox 2001; Jasperson 2002; Local & Walker 2005; Curl 2002, 2004; Curl et al. 2004; Walker 2004a, 2004b; Ogden et al. 2004). These studies have demonstrated that rather than spoken language needing to be *decontextualised* in

order for its variability to be constrained, fruitful insights can instead be had if it is *contextualised* in its interactionally-embedded context of production. This is because it is within the embedded context of production itself that these studies have shown many aspects of language to receive: (1) their constraints, as evidenced by the sequentially determined, systematic and recurrent patterns of phonetic detail observed; and (2) their meaning, as evidenced by the orientations of the participants to the sequential details of the interaction.

This thesis contributes to this expanding area of research which integrates CA and phonetic analysis to explore the relationship between the phonetic and the interactional structure of naturally-occurring talk-in-interaction. It aims to identify whether there are any orderly and recurrent mappings between phonetic-linguistic forms and interactional functions/meanings. A fundamental assumption in this thesis is that language receives its meaning primarily from its context of production in talk-in-interaction, and that meaning and context are therefore inseparable: an assumption taken from the Firthian approach to language and borne out in this research (Firth 1935a). To uphold such a view necessitates a particular take on language meaning - one that regards meaning as being context-bound, emergent and contextually determined which, in turn, obliges a polysystemic view of language. These beliefs, which again have their origin in the Firthian approach and are embodied in the methodological approach of CA, comprise the foundations of this research. They are markedly different from those held in many current linguistic-phonetic investigations which are, instead, concerned with ascribing context-free lexical and propositional meanings to language.

Central to this thesis are the notions of *variability*, *context* and *meaning*. The first chapter will outline some of the key and influential approaches which have embraced these notions and which are of relevance to this research. These approaches comprise the Firthian Approach, the Variationist Sociolinguistic Approach and Conversation Analysis.

CHAPTER ONE

CONTEXT

1.1 Introduction

This thesis focuses upon the interface between the linguistic-phonetic and the interactional-sequential design of English conversation. In particular, it examines the relationship between clicks and their embedded contexts of production in the sequential and interactional structure of talk. This chapter provides the intellectual context from which the research in this thesis arose, situating the thesis within the academic arena. Three influential approaches are discussed which concern the notions of variability, context and meaning - three concepts that are central to this study. First, the Firthian approach is presented (1.2), in which the internal, embedded context of language production is regarded as being of paramount importance to language meaning. Next, the Variationist Sociolinguistic approach is outlined (1.3) which, contra and in response to the generative decontextualisation of language, places variability at the heart of an analysis. Afterwards, the Conversation Analytic approach is discussed (1.4) and a direct comparison is drawn between the Firthian approach and the methodological framework of Conversation Analysis (henceforth CA). This comparison demonstrates that these two approaches share many similar ideas about language meaning and context. Subsequently, a direct comparison between the Variationist *external* and the CA *internal* invocations of context is presented and the differences between these two approaches (Firthian and CA) are exemplified using a data fragment (1.5). Finally, the directions of and motivations for the study are highlighted (1.6) and the organisation of the thesis is outlined (1.7).

1.2 The Firthian Approach

The notion of meaning featured highly in the work of J. R. Firth, as he ardently believed that ‘the object of linguistic analysis ... is to make statements of meaning so that we may see how we use language to live’ (1957: 192). There are three key and fundamental aspects of the Firthian concept of meaning which are of relevance to the research in this thesis and will therefore be discussed in this chapter. The first is Firth’s notion of Context of Situation, which is of particular import, as Firth regards all meaning in talk as being contextually determined and hence argues that ‘no study of meaning apart from a complete context can be taken seriously’ (1935a: 7). The second central property of Firth’s take on meaning is his proposal that meaning is distributed across many levels of analysis and that statements about meaning should therefore incorporate all of these levels, and thus produce a ‘spectrum of statements’ (rather than just focusing on one exclusive level). The third key feature of Firth’s approach to meaning, which is of interest to this research, is his eschewal of appeals to non-observable psychological phenomena. Instead, Firth insists that investigations of meaning must be contextually embedded and grounded in the actual observable conduct of the parties-to-talk. These three fundamental features of the Firthian approach—Context of Situation, spectrum of statements and rejection of psychological phenomena—are discussed in 1.2.1 to 1.2.3 below respectively.

1.2.1 Context of Situation

A defining feature of Firth’s views on meaning is the notion of Context of Situation (see Kelly & Plug (eds.) forthcoming, for a discussion of the Firthian approach). This concept was taken from the anthropologist Malinowski (1923), who undertook much research in preliterate societies in the South Seas.¹ Although Malinowski could

¹ See Firth (1950) for differences in the use of the term ‘context of situation’ between Firth and Malinowski.

converse freely with the Islanders, he experienced many difficulties when attempting to translate from their native language into English, as the two cultures in which these languages were used were so different. In order to address and combat these difficulties, Malinowski became concerned with establishing a theory of meaning and language use. He proposed that ‘the meaning of any utterance is what it *does* in some context of situation’ (Dinneen 1967: 299; emphasis added). Malinowski thus coined the term ‘Context of Situation’ to mean ‘the environment of the text’; that is the verbal and the situational environment of the language production (Halliday & Hasan 1985: 6). In the words of Malinowski himself:

‘utterance and situation are bound up inextricably with each other and the context of situation is indispensable for the understanding of the words...a word without *linguistic context* is a mere figment that stands for nothing by itself, so in the reality of a spoken living tongue, *the utterance has no meaning except in the context of situation*’ (1923: 307; some emphasis added).

Unlike Malinowski who was ‘forced into linguistics by the needs of his own subject...Firth was a linguist, deeply concerned with linguistic theory, who was forced into ethnography in order to achieve an adequate understanding of meaning’ (Robins 1971: 36). Following Malinowski, Firth also proposed that the meaning of language can only be understood by examining language as a contextualised production, i.e. in its context of situation, and by investigating what its function is in the context in which it is produced. In fact, Firth considered the context in which talk is embedded to be so crucial to its meaning that he argued that: ‘Each word when used in a new context is a new word’ (1951: 190).

Although Context of Situation was a key feature of the Firthian approach, it was a notion, like many others, which was not fully developed or defined. In one of his latest and most comprehensive works, *A Synopsis of Linguistic Theory*, Firth (1957: 177)

acknowledged the lack of technical description of Context of Situation and the potential for further research in this area:

‘The technical language necessary for the description of contexts of situation is not developed, nor is there any agreed method of classification. At this level there are great possibilities for research and experiment’.

In this paper, Firth (ibid: 177) also presented his clearest definition of (his concept of) a Context of Situation by proposing the relevance of the following categories:

1. The participants: persons, personalities and relevant features of these.
 - (a) The verbal action of the participants.
 - (b) The non-verbal action of the participants.
2. The relevant objects and non-verbal and non-personal events.
3. The effect of the verbal action.

Thus the main components of Firth’s schematic construct of the Context of Situation are ‘the human participant or participants, what they say, and what is going on’, and it is the interrelation between these main elements that provides the full Context of Situation (Firth 1935a: 27). Aside from this general schema, however, Firth did not offer any further details or exemplifications of his concept of Context of Situation. This gap was noted by Robins (1971: 37): ‘It is notorious that Firth provided totally inadequate exemplification of a semantic analysis in terms of the theory he so ardently believed in and so assiduously preached’. The lack of concrete definitions in Firth’s approach was intentional, however, as evidenced in the following quote: ‘I have nothing to say on the standardization of terminology nor do I consider nomenclature in itself of primary importance’ (Firth 1956a: 120).

Perhaps the clearest and most obvious application of Firth's notion of Context of Situation is presented in a study by Mitchell (1957). In this study, Mitchell examines the (Arabic) language used in various stages of one type of situated activity: the activity of buying and selling in Cyrenaican markets.² In line with Firth's Context of Situation schema, Mitchell situates his analysis in its context of production and includes factors such as the features of the participants. He states:

'By adhering to the principle that meaning must be sought in use, we are able at the situational level to make a systematic classification of material on the basis of correlations between texts and their environments' (Mitchell 1957: 32).

Although the Firthian notion of Context of Situation was not fully worked out or exemplified, it was nonetheless one of the defining components of the Firthian approach. Firth (1950) argues that the Context of Situation 'makes sure of the sociological component' (p. 182), as it is an 'abstraction at the social level of analysis' (p. 183). In the Firthian school of thought, then, the context in which language is produced is considered to be inseparable from its meaning and its meaning to be inseparable from its context.

1.2.2 Multi-Level Statements of Meaning

Recall that Firth regards 'the object of linguistic analysis' to be that of making 'statements of meaning so that we may see how we use language to live' (1957: 192). According to Firth (*ibid*: 173f), these statements should be produced at different levels of analysis, as he proposed that 'the meaning of [language] ... cannot be achieved at one fell swoop by one analysis at one level'. Instead, Firth regarded meaning as being 'a complex of contextual relations, and phonetics, grammar, lexicography, and semantics each handles its own components of the complex in its appropriate context' (1935a: 19).

² See Settineri (1998) for a modern-day analysis of the language of buying and selling in Sicily.

Statements about and analyses of meaning should therefore, in Firth's view, be 'dispersed in a spectrum of specialized statements' (1950: 183) (which he likens to the dispersion of light in different wave-lengths in a spectrum) and tied to their situated contexts of production.

In 1950, Firth claimed that Context of Situation can be used as 'the basis of the hierarchy of techniques for the statements of meanings' (p. 183). However, one year later he modified this claim by arguing that statements of meaning can begin *either* from the context and work down, i.e. a top-down approach, or from the phonetics and work up, i.e. a bottom-up approach:

'To make statements of meaning in terms of linguistics, we may accept the language event as a whole and then deal with it at various levels, sometimes in a descending order, beginning with social context and proceeding through syntax and vocabulary to phonology and even phonetics, and at other times in the opposite order' (Firth 1951: 192).

Again, however, Firth changed his line of argument by subsequently claiming that one should begin one's investigations of meaning with context:

'The meaning of texts is dealt with by a dispersal of analysis at mutually congruent series of levels, *beginning with contexts of situation* and proceeding through *collocation*, syntax (including *colligation*) to phonology and phonetics'

(Firth 1957: 200f; some emphasis added).

It is clear from the differing viewpoints in the above quotes that in the Firthian approach, the starting point of an investigation—whether one conducts a top-down or a bottom-up exploration—is not important. What is important, however, is that the investigator does not focus on just one level of language, such as syntax or phonology, but instead examines *all* levels of analysis, *including the social context*, as it is within

each of these levels that meaning is regarded as being situated. Moreover, each of these levels must also be relatable to each other by an ‘analysis at a series of congruent levels’ (Firth 1957: 199), and then be ‘synthesis[ed] on renewal of connection in experience’ (ibid: 177). In other words, the Firthian approach maintains that although analyses of meaning should be distributed across different levels in a ‘spectrum’ method of analysis, these levels should subsequently be interlinked once again and related to other language material produced in context by interactants in the real world.

1.2.3 Rejection of Psychological Phenomena

The third and final essential feature of Firth’s approach to meaning which is of particular relevance to this thesis is his eschewal of the idea that meaning is concerned with ‘relations in a hidden mental process’ (1935a: 19). Firth’s rejection of hidden psychological factors and mentalist entities being a necessary component of linguistic statements of meaning was radical for his day, as many of his predecessors and contemporaries took the opposite dualist stance: they advocated a split between mind and body whereby ‘form and meaning must be correlated in some matching process’ (Read 1955: 39; see ibid for an overview of some of the problems with the use of the term ‘meaning’ in the twentieth century). One example of this dualist approach is presented by Ogden and Richards (1923) in their book *The Meaning of Meaning*. In this work these scholars opine that: ‘The analysis of the process of communication is partly psychological’ (ibid: 8). Thus they argue for a three-point relational view of meaning comprised of symbols (i.e. the words), referents (i.e. the events) and *thoughts*, the latter of which is the link between the symbols and referents (ibid: 11; emphasis added).

In contrast to the dualist approach, Firth claims there is ‘no point trying to discover the essence of meaning either in the intention of humans or by analysing thought processes’ (1952a: 181). He argues:

‘As we know so little about mind and as our study is *essentially social*, I shall cease to respect the duality of mind and body, thought and word, and be satisfied with the whole man, thinking and acting as a whole, in association with his fellows’ (1935a: 19; emphasis added).

Firth therefore proposes a ‘contextual approach’ to meaning, which concerns ‘the interrelations of the terms of the *actual observable context itself*’ (Firth 1935a: 19; emphasis added). He states:

‘the contextual approach, while not concerning itself either with the inclusion or exclusion of “mentalist” factors, regards linguistic action as generally meaningful in the sense that the maintenance of the patterns and processes of our common lives is regarded as meaningful activity, and the systemic and structural analyses of such activity, as studies of meaning at a series of levels’ (Firth 1952a: 181).

In other words, Firth regards meaning in language as being socially constructed and contextually embedded, and as being established from observations about what the participants are doing in, with and through their talk rather than from speculations about what is going on in their minds.

1.2.4 Summary

In this section (1.2) the Firthian approach to language has been examined. Unlike many of its contemporaneous approaches, this approach regards statements of meaning as being the central goal of linguistic analysis. Moreover, it regards all meaning as being contextually determined. Psychological explanations of meaning which are not grounded in the observable conduct of the participants are therefore eschewed. In addition, meaning is considered to be an inherent property of all levels of analysis. This approach therefore insists that statements of meaning should be conducted at a series of congruent levels.

One of the key features which percolates throughout the Firthian approach is the firm belief that in order to understand the workings of language, the analyst must focus on talk which is produced and embedded within its context of production. However, although there was great interest in the analysis of talk housed within its context of production in the early part of the twentieth century, much of this interest petered out soon after the publication of Chomsky's (1957) seminal and highly influential work (and Firth's death in 1960 (Robins 1971: 33)).

In the traditional generative tradition (founded by Chomsky 1957, 1965), variability and context were typically removed from investigation, as they were thought to tarnish and obscure the important underlying invariance. Chomsky's work gave rise to a new direction in linguistic theory which focused on a person's competence and the formal aspects of language rather than their performance and the contextualised components of language use. The generative exclusion of variability and context greatly influenced linguistic-phonetic research and has resulted in a general overemphasis on a speaker's 'competence'. Such decontextualisation of language from its embedded interactional environment is surprising given that, as stated by Abercrombie (1965: 3), conversation represents 'the most natural, the most frequent, and the most widespread occurrences of spoken language'. However, although the traditional, generative approach has been highly influential in linguistic-phonetic theory, not all linguists have ascribed to its methods of investigation. Instead, some linguists have embraced the notions of variability and context and placed them at the very core of their investigations. One such notable exception is the Variationist approach, pioneered by Labov in the 1960s. This approach was highly radical for its time, as it strongly opposes the Chomskyian decontextualisation of language and the dismissal of variability as being unstructured. Instead, the Variationists concern themselves with identifying and accounting for socially determined patterns of linguistic-phonetic variability in socially situated speech productions. The next section (1.3) will examine the Variationist approach.

In Variationist investigations, internal factors are always the first to be examined, as it is well known that some phonetic variation in talk can be conditioned by system-internal linguistic constraints. For instance, factors such as phonetic and phonological context (Wells 1982; Mees 1990; Eckert 2000), grammatical status (Guy 1980; Sankoff 1980) and syllable structure (Labov 1994; Milroy 1992) have all been previously shown to influence the frequency with which particular variants may occur. However, although the identification of the system-internal linguistic constraints is an important and informative part of the Variationist analysis, it cannot (and is by no means used to) account for all the patterns of variation observed in socially situated speech. Rather, in order to account for the linguistic-phonetic patterns which are not constrained by the linguistic system, Variationists seek to relate the variability found to *external contextual factors* of the speech production.

The most typical external contextual features invoked in Variationist accounts of talk comprise macro level social characteristics of the participants, such as their age, sex or social class.³ These contextual features will be examined in 1.3.1. Afterwards, two key methodological principles which underpin Variationist research will be discussed: the notion of the linguistic variable (1.3.2) and quantification (1.3.3).

1.3.1 Macro Contextual Feature: Social Attributes of Speaker

The defining feature of Variationist sociolinguistics is the belief that variation in talk is ‘socially conditioned’ and that the social characteristics of a speaker determine their linguistic choices (Downes 1998: 98). Variationists therefore seek to account for variation in socially embedded speech productions by identifying the relationship between certain linguistic variables and an individual’s social attributes, such as their

³ Additional external contextual features commonly appealed to in Variationist studies are the stylistic features of the speech production (see e.g. Labov 1971, 1972a; Wolfram & Fasold 1974; Trudgill 1974; Mathisen 1999; Tollfree 1999) and the social networks of the participants (see e.g. Douglas-Cowie 1978; Milroy 1980, 1987; Eckert 2000; Milroy & Gordon 2003).

age, gender or social class. A speaker's social attributes are considered to be the 'independent' variables whereas the linguistic variables are considered to be the 'dependent' variables, as the latter are regarded as being dependent on the social characteristics of the speaker. The key empirical goal in Variationist studies is to correlate a speaker's linguistic-phonetic (dependent) variables with their social (independent) variables (Chambers 1995: 17).

Macro-investigations of group-correlated variability have long been the main focus of much Variationist research and numerous scholars have argued that there is an intimate relationship between the linguistic-phonetic patterns of variability in socially embedded talk and the social, non-linguistic attributes of the speaker (see e.g. Trudgill 1972, 1974; Macaulay 1977; Coates 1986; Labov 1990; Milroy & Milroy 1991; Docherty & Foulkes 1999).⁴ In order to examine this relationship, Variationists are typically highly selective with their speech data, taking it from a 'representative sample of the performance of a defined community' (Docherty 2003: 11). To obtain their representative sample Variationists usually employ a 'quota/judgment sampling technique', rather than a 'random sampling procedure' (Milroy & Gordon 2003: 30). This means they firstly identify the types or categories of speakers to examine after which they seek the speakers needed to fit their pre-determined categories (Milroy 1987: 26).

The categories of speakers examined in Variationist studies are typically stratified according to their social characteristics (see Sankoff 1980 for typical sampling procedures). For example, the Variationist researcher may use a stratified sample consisting of individuals from four different age groups, two genders and four social classes. If the sample comprised four speakers per cell, as is typically taken to be the minimal number in Variationist studies (see e.g. Watt & Milroy 1999, Docherty &

⁴ But see Milroy & Gordon (2003: 116; emphasis added) for a criticism of the Variationist focus on the group rather than the individual; these scholars argue that 'categories such as class, gender and ethnicity are *macro-level analyst constructs*' and that the most important factor in determining a person's linguistic choices may be a person's *micro* social networks.

Foulkes 1999), this would give a sample size of 128 ($4 \times 4 \times 2 \times 4 = 128$). Obviously, of sample size would create a vast amount of data, which could be time consuming to examine. Smaller samples are therefore often used but have the added disadvantage of the generalisations made about language being based on small numbers of speakers (Milroy 1987: 22). Many studies therefore restrict the social stratifications by having, for example, only two age groups and two social classes (see e.g. Watt & Milroy 1999, Docherty & Foulkes 1999). Such a stratification (which included two genders) would comprise a sample size of 32, should the typical four speakers per cell be used ($4 \times 2 \times 2 \times 2 = 32$).

Overall, the most studied social characteristics in Variationist analyses have been age, class and gender (see e.g. Wolfram 1969; Trudgill 1972, 1974, 1986, 1988; Macaulay 1977; Coates 1986; Coates (ed.) 1998; Eckert 1989, Labov 1972b, 1990; L. Milroy & J. Milroy 1991; J. Milroy & L. Milroy 1998; Docherty & Foulkes 1999; Eckert 1998; Mees 1987; Mees & Collins 1999).⁵ With regard to age and gender, this is perhaps due, in part, to these two independent variables being relatively easy to relate to a person's linguistic variables (and their variants). The influence of a person's social class on their linguistic choices, on the other hand, is much more difficult to explore, as although numerous scholars have suggested various ways of defining and measuring a person's class (see e.g. Labov 1966a; Trudgill 1974; Macaulay 1976), there remains no clear definition as to what social class *is* nor how a person's class can be *measured* (see Milroy & Gordon 2003: 95-100 for a discussion of some of the problems encountered when attempting to define and measure social class). Typical factors appealed to when classifying a person's social class include occupation, education, income, housing and, more recently, lifestyle (Wardhaugh 1992: 145). Even though its definition remains 'inherently fuzzy', social class is still considered to be one of *the* most important independent variables in Variationist research. According to Chambers (1995: 36), our

⁵ See Eckert (1989) for a discussion on the use of the terms *sex* and *gender*.

social class comprises ‘the most linguistically marked aspect of our social being’. It has consequently received much attention, particularly in earlier sociolinguistic investigations (see e.g. Labov 1966a, 1966b, 1972c; Wolfram 1969, 1993; Fasold 1972; Trudgill 1974; Reid 1977; Rickford 1986; Chambers & Trudgill 1980). Further discussions on the notions of class, age and gender and the interrelationship between these variables lie outside the scope of this chapter. The reader is referred to the references given above for more information.

1.3.2 The Linguistic Variable

A fundamental component of Variationist research is the concept of ‘the linguistic variable as a structural unit’ (Milroy & Gordon 2003: 5; see Labov 1966b, 1972c and Chambers 1995 for a full account of this concept). The linguistic variable, according to Chambers and Trudgill (1980: 60), is ‘a linguistic unit with two or more variants involved in covariation with other social and/or linguistic variables’. The variables all share in common the fact that they belong to a ‘linguistically defined set’ (Wolfram 1993: 195), although the sets can belong to numerous different levels of organisation such as the phonology, the syntax or the morphology. If a linguistic variable can be correlated with an extra-linguistic social variable, such as a person’s age or sex, then it is regarded as being a ‘sociolinguistic variable’ (Downes 1998: 96).

Variationists have examined a wide range of phonological variables. The most widely studied, particularly in North America, have been vocalic and following Labov et al’s (1972; cited in Docherty 2003: 12) seminal work, instrumental phonetic techniques have been extensively employed to investigate vocalic variables (see e.g. Labov 1994; Clark et al. 1995; Watt & Tillotson 1999). Far less studied are consonantal variables and although there has been a fair amount of auditory phonetic investigations of consonants, there has, until recently, been no instrumental phonetic analyses of consonantal variability in socially situated speech productions. This gap in knowledge is currently

being addressed by a number of Variationists (or perhaps more appropriately, Sociophoneticians) in Great Britain (see e.g. Docherty et al. 1997; Docherty & Foulkes 1999; Foulkes & Docherty 2000). These scholars have sought to ‘look at what value could be added to our understanding of sociolinguistic variation by bringing to bear a more refined (auditory and instrumental) phonetic analysis than is typically employed in the analysis of consonantal variation’ (Docherty & Foulkes 2000: 105). They have argued that the deployment of sophisticated instrumental techniques enables fine-grained systematic patterns of consonantal variation to be identified (see e.g. Docherty & Foulkes 1999; Docherty 2000). Far fewer studies have focused on other aspects of phonological variation such as ‘intonation’ (but see Pellowe & Jones 1978; Britain 1992; Grabe et al. 2000) and voice quality (but see Stuart-Smith 1999; Stuart-Smith & Lawson 1999).

An important quality of the (socio)linguistic variable is that it does *not* function categorically whereby speakers either do or do not use certain variables. Instead, it enables statements to be made about the greater or lesser frequency of occurrence of particular variants (in particular situations). The concept of the linguistic variable therefore does away with the axiom of categoricity (i.e. ‘the simplifying assumption that data for linguistic analysis must be regularized to eliminate real-world variability’ (Chambers 1995: 12)) and instead allows direct analyses and systematic comparisons of vast amounts of naturally-occurring speech data (Milroy 1987: 94; Milroy & Gordon 2003: 5).

1.3.3 Quantification

An additional key methodological feature of Variationist studies is the deployment of statistical measures of significance to assess the importance of an observation. The methodology typically adopted in these studies is ‘to identify a candidate sociolinguistic variable and *to quantify all of the variants observed* within a representative sample of

the performance of a defined community' (Docherty 2003: 11; emphasis added). There are two main ways in which the variants can be quantified, according to which *type* of variants are identified in the analysis: if the variants are *discrete*, as in the presence or absence of (r) in Labov's (1966a) study of New York English, then a simple percentage score can be used, which denotes the percentage occurrence or non-occurrence of a variant. If, on the other hand, the variants are *continuous* (i.e. a choice between more than two variants along a phonetic continuum), as found for (t) in Docherty et al. (1997), for example, and is typically the case with vowels, then an index score (a mean score of the speaker's usage) is used rather than a percentage.

The deployment of an index score involves arranging the variants along a single phonetic dimension and then assigning each variant with a social value according to its degree of standardness: the most standard forms get a score of zero and the less standard forms score more highly (but see Milroy 1987 for a discussion of some of the criticisms regarding the assumption that the variants of a sociolinguistic variable can be placed on a single phonetic and social dimension). For the sake of brevity, the continuous variables 'may be treated as having discrete variants' (Milroy & Gordon 2003: 139). One study in which continuous variables were 'converted' into discrete variants is that by Foulkes & Docherty (2000) in which the realizations of (r) occurred on a continuum from [ɹ] to [v] but an arbitrary four point scale was used to break the continuum: the [ɹ] variant was assigned a score of one and the [v] variant a score of four with two variants in between having scores of two and three according to if they sounded more like [ɹ] or [v]. The resulting scores, whether presented as binary percentage scores or as weighted index scores, enable easy comparisons to be drawn across a vast amount of data and speakers (Williams 1992: 73).

After the distribution of the variants has been established, Variationists typically establish the significance of their findings by applying statistical tests. The purpose of

these tests is to obtain independent evidence either to support or refute a given hypothesis about linguistic variation and, as stated by Fasold (1984: 109), ‘to distinguish valid numerical differences from those that are due to mere random factors’. There are many statistical techniques that can be applied and their application depends on the type of data being examined. Typical tests used include t-tests, correlations and the sociolinguistic software packages Goldvarb™ and Varbrul™ (see Fasold 1984: 85-112 for more information on statistical tests).

1.3.4 Summary

In this section the Variationist approach has been outlined. This approach, contra to the traditional generative approach, places variability at the very heart of an investigation. In seeking to account for variability in talk, Variationists typically appeal to contextual factors *external* to the speech production, such as the participant’s age or sex, as a speaker’s social attributes are regarded as determining their linguistic choices. This approach is therefore markedly different from the Firthian approach in which all meaning in language use is regarded as being bound to its context of production rather than to external sources. Two key methodological principles which underpin Variationist research have also been presented: the concept of the linguistic variable and the deployment of quantification measures.

The Variationist approach was influential in that it readjusted the focus of linguistic-phonetic research from an over-emphasis on decontextualised speech productions to one that includes investigations of socially-situated speech productions. It has not, however, explored the linguistic-phonetic properties of the organisation of actual, naturally-occurring conversation *per se*. Instead, it has been primarily concerned with identifying and accounting for socially determined patterns of linguistic-phonetic variability. To date, the linguistic-phonetic properties of the organisation of naturally-occurring talk-in-interaction are little studied compared with the myriad of research into the phonetic

properties of carefully controlled lab or intuited data (but see e.g. Shriberg 1994, 2001; Simpson 1991, 1992a, 1992b, 2001; Douglas-Cowie & Cowie 1998; Jurafsky et al. 1998; Jurafsky et al. 2001; Couper-Kuhlen & Selting (eds.) 1996; Kohler 2000, 2001; Wesener 2001; Ford et al. (eds.) 2002; Shriberg et al. 1998; Bell et al. 2003; Couper-Kuhlen & Ford (eds.) 2004).

Although the *linguistic* characteristics of conversation have been little studied, the *interactional* properties of actual, ordinary everyday talk-in-interaction have received much attention from scholars within the sociological field of Conversation Analysis⁶ (see e.g. Sacks et al. 1974; Schegloff et al. 1977; Schegloff 1979a, 1979b, 1986a, 1986b, 1987b, 1995a; Jefferson 1974, 1980a, 1980b, 1984a, 1984b; Sacks 1987, 1992a, 1992b). The Firthian ideas about language meaning and context (see 1.2 above) are strikingly similar to those held in Conversation Analysis (CA) even though these fields were established centuries apart (the former being started in the 1930s and the latter in the 1960s), and technically comprise different academic disciplines. Recent years have seen a number of linguistic studies using CA as a methodological toolbox with which to study (many of Firth's ideas about) language (see e.g. Local et al. 1985; Local & Kelly 1986; Local et al. 1986; Local 1992, 1996, 2004; Couper-Kuhlen 1993, 2001a, 2004a, 2004b; Tarplee 1993; Wells & Peppé 1996; Wells & Macfarlane 1998; Ogden 2001, 2002, 2003, 2004; Wright 2001, 2002a, 2004; Local & Walker 2005; Curl 2002, 2004, 2005; Curl et al. 2004; Walker 2003, 2004a, 2004b; Ogden et al. 2004). These studies have successfully demonstrated that, as originally claimed by Firth and his co-workers, various patterns and regularities can be observed in the linguistic-phonetic properties of talk-in-interaction by exploring language embedded in its natural habitat of social interaction.

⁶ By here separating the levels of linguistics and interaction, the author is not wishing to advocate that analyses should focus on one or the other of these levels of organisation nor, in fact, that they are independently organised.

In this thesis the micro-oriented, data-driven methodological approach of CA is employed to investigate the linguistic-phonetic properties of talk-in-interaction from a Firthian perspective. The next section (1.4) will therefore examine CA and its applicability as a Firthian toolbox by drawing a number of direct comparisons between CA and the Firthian approach.

1.4 Conversation Analysis (CA): A Firthian Toolbox

‘Consider...the philologist or historical linguist of the distant or proximate future who treats as the linguistic remains of contemporary society not scrolls, books, or memoranda but film and video/audio tape of everyday, spontaneous interaction in the lives people live. Imagine as well that such a linguist is not committed to a theoretical set and to terms of analysis like those currently familiar but is prepared to derive the appropriate terms of analysis from the materials under investigation. What understanding of the English language might result if not only the analyses but also the very terms of analysis were formed on the basis of such materials?’

(Schegloff 1979a: 263).

CA is a rigorously empirical, qualitative approach to the study of the ‘sequential organisation of interaction’ (Heritage 1989: 24). As eloquently stated by Ford et al. (2002: 4):

‘For more than thirty years, CA has offered ... a theoretical and methodological foundation for documenting the interactional construction of social orders...[where social order] is understood as practice, an order created by participants in talk-in-interaction—jointly, contingently, and always locally...[CA was developed] to account in a detailed, data-driven manner for the resources and practices used in creating social orders through situated language use’.

CA therefore presents the micro-oriented, language analyst with a sophisticated and principled way of examining and describing the ways in which participants-in-talk devise, manage and orient to the sequentially unfurling interactional structure of conversation (for an overview of CA see Heritage & Atkinson 1984; Heritage 1985, 1989; Zimmerman 1988; Hutchby & Wooffitt 1998; ten Have 1999; Drew 2005).

It is important to point out that, as outlined by Drew (2005: 73), the term ‘Conversation Analysis’ is a ‘misnomer’ for two reasons. First, it originated from a more general inquiry ‘concerning whether a stable, reproducible, cumulative, natural observational science of social action, and hence of society was possible, and if so, how it could be conceived’ (Drew 2005: 73). Conversation was thus examined not for its own sake but because it was one accessible form of conduct. Second, the focus of CA is not only conversation but rather any form of social interaction in which language is used. This includes talk in a courtroom between a defendant and a judge, talk in a doctor’s surgery between a patient and a doctor and talk in a classroom between a teacher and a student. Such talk is not a conversation yet the tools of CA extend to include analysis of it. CA can therefore be regarded as a ‘method of analysis’ (Drew 2005: 73), which can be applied to all talk produced in interaction. Despite being a misnomer, the term CA has remained and will therefore be used in this thesis. The term ‘talk-in-interaction’ (Schegloff 1987b: 207) will also henceforth be used rather than *conversation* to reflect the wider analytical scope of CA and the fact that talk is typically interactionally embedded, as is the case in the data employed in this thesis (see chapter 2 for details of the data).

CA and the Firthian school of thought share many striking similarities in their views on and approaches to language. This section will outline five of these in the following sections: Talk-in-Interaction as Social Action (1.4.1); Actions are Context-Shaped and Context-Renewing (1.4.2); Polysystemicity (1.4.3); Rejection of Intuited and

Experimental Speech Data (1.4.4) and Participant Orientation as a Basis for Theorising (1.4.5).

1.4.1 Talk-in-Interaction as Social Action

Both CA and the Firthian approach place great emphasis on examining speech as a contextualised production embedded in its natural habitat of social interaction. This was stressed by Firth in 1935:

‘Neither linguists nor psychologists have begun the study of conversation, but it is here we shall find the key to a better understanding of what language really is and how it works’ (1935a: 32).

The importance Firth placed on contextually bound analyses of talk ‘immersed in the immediacy of social intercourse’ (Firth 1952b: 13) stemmed from his regard of language as being a ‘social instrument’ (Firth 1937: 15) which is used as a ‘mode of action in a context of situation’ (Firth 1930: 14). In fact, Firth considered ‘the disturbance of the air and other people’s ears by means of bodily utterance’ (1937: 19) to be *the* most important social action undertaken by human beings. In a similar vein, Conversation Analysts also insist that investigations of language should be based on ‘naturally-occurring occasions of everyday interaction’ (Heritage & Atkinson 1984: 2). This is because, like Firth, they regard talk-in-interaction as being ‘a/the primordial site of human sociality and social life’ (Schegloff 1987a: 101) and therefore argue that ‘conversational interaction... is sociological bedrock’ (Schegloff 1995a: 187).

1.4.2 Actions are Context-Shaped and Context-Renewing

CA and the Firthian approach both maintain that in order to identify and understand the actions that parties-to-talk produce and orient to in talk-in-interaction, it is essential that one examines talk embedded in the context in which it is produced. Context thus plays

an equally essential role in these approaches, as evidenced in the following quotes produced by Firthian and CA scholars respectively:

‘no study of meaning apart from a complete context can be taken seriously’ (Firth 1935a: 7).

‘any participant’s communicative action is...context-shaped because its contribution to an ongoing sequence of actions cannot be adequately understood except by reference to the context in which it participates’ (Heritage 1989: 22).

Importantly, in both the CA and Firthian traditions, context is not regarded as being a static entity. Instead it is considered to be continually reshaped and renewed as the language sequentially unfolds, which is evident from the following quotes produced by Firthian and CA scholars respectively:

‘In any context of situation, the normal human being and his environment are one; the past merges in the present in which the future is always on the point of being born’ (Firth 1937: 20).

‘Since every ‘current’ action will itself form the immediate context for some ‘next’ action in a sequence, it will inevitably contribute to ... [how] the next action will be understood. In this sense, the context of a next action is repeatedly renewed with every current action’

(Heritage 1984a: 242).

In other words, CA and the Firthian approach argue that at each point in talk, the current context presents an understanding of the preceding context whilst simultaneously renewing the context for the next action. To use Heritage’s words: ‘the significance of any participant’s communicative action is *doubly contextual*, in that the action is both context-shaped and context-renewing’ (1989: 22; emphasis added).

If one accepts the notion that language meaning is always context dependent, then, according to CA and the Firthian approach, potential ambiguities in talk, such as that found in the phrase *who'd you give it to?* (Firth 1930: 43), are not relevant or problematic. This is because these potentially ambiguous stretches are not produced in isolation in talk-in-interaction but rather are embedded within a specific context of production which provides (and creates) their interpretation (Schegloff 1984: 51). The CA and Firthian context-embedded approach to language analysis therefore differs markedly from most linguistic approaches to talk and meaning, as linguists typically analyse language divorced from its situated empirical context, thus giving analytic supremacy to isolated stretches of talk (see e.g. Speech Act Theory: Austin 1962; Searle 1969). In comparison, CA and Firthian scholars maintain respectively that 'utterances ... are contextually understood by reference to their placement and participation within sequences of actions' (Heritage & Atkinson 1984: 5); and that we must therefore 'take our facts from speech sequences, verbally complete in themselves and operating in contexts of situation which are typical, recurrent and repeatedly observable' (Firth 1935b: 35).

1.4.3 Polysystemicity

According to CA and the Firthian approach, meaning in talk can only be established by examining speech embedded within its context of production in talk-in-interaction. This therefore means that the sequential placement of talk is regarded as providing its interpretation, resulting in two interconnected outcomes: (1) identical components of language use, such as lexical items or phrases, can have different interpretations according to their sequential location, as in different interactional environments they can be used to undertake different interactional work; (2) identical interactional functions can be undertaken by a range of different linguistic objects. To align with these two views necessitates a polysystemic approach to language.

The belief that language is polysystemic and should therefore be analysed as being comprised of a ‘system of systems’ (Henderson 1987: 63) is an additional similarity between CA and the Firthian approach. Polysystemicity featured highly in the work of Firth and his practitioners, as they believed that a polysystemic approach to language was necessary in order to gain insights into the complexities of its use (see e.g. Henderson 1949, 1951; Carnochan et al. 1954; Sprigg 1957; Carnochan 1957, 1960).⁷ Contra to arguments that polysystemic investigations are unnecessarily complicated, the Firthians argued that the ‘guiding spirit of polysystemic analysis...is not so much a desire for complexity for its own sake as a dislike of inaccurate oversimplification and specious simplicity’ (Sharp 1958: 128), features which the Firthians regarded as being prevalent in the predominant ‘monophysite’ doctrines of language (Firth 1948: 123).

In comparison, while CA scholars have not overtly stated their belief in a polysystemic approach to language it exudes from the analytic and methodological standpoints which underpin the CA approach. Work which has adopted a CA-oriented methodology has evidenced, for example, that various lexical items such as *and* (Wright 2001, 2002a; Local 2004), *actually* (Clift 2001) and *never mind* (Couper-Kuhlen 2004a) can undertake different interactional activities according to their sequential placement. These items therefore comprise terms in different systems which are structurally demarcated and set up on an *ad hoc* basis (Firth 1948; Heritage 1984a) according to interactional function.

⁷ See Ogden & Kelly (2003) for a discussion on the Carnochan et al. (1954) analysis. See also Simpson (1992a) and Ogden (1999) for modern day polysystemic accounts of the English auxiliary system.

1.4.4 Rejection of Intuited and Experimental Speech Data

The CA and Firthian insistence that investigations of language should be empirically grounded and contextually embedded results in both of these approaches rejecting the use of intuited or laboratory controlled speech data. This rejection immediately sets these traditions firmly apart from the majority of linguistic (in particular phonetic and phonological) approaches, as many linguists typically regularise variability in speech by working on either intuited or experimental data (see e.g. Lehiste 1975; Crystal & House 1990; Selkirk 1984, 1995; Nespor & Vogel 1986). In comparison, ‘conversation analysts have approached their topics by treating the ‘real world’ as their laboratory (Heritage 1989: 24), as they believe that: ‘In any environment of so-called language use, there is a locally organised world in which it is embedded’ (Schegloff 1979a: 282). This view resonates strongly with Firth’s comments about taking ‘linguistics into the laboratory rather than look[ing] into laboratories for linguistics’ (Firth 1957: 202), by which he means that one should examine actual naturally-occurring talk-in-interaction (embedded in its context of production) for patterns or regularities rather than search in contrived data for evidence of a pattern or regularity which may be nothing more than an artefact of the way in which the data were constructed.

One reason for the CA and Firthian rejection of experimental speech data is that, as argued by Heritage (1984a: 237), one cannot intuit the range of contexts invoked and oriented to in conversation:

‘The myriad ways in which specific contexts (e.g. particular social identities, purposes and circumstances) are talked into being and oriented to in interaction vastly exceed the comparatively limited, and overwhelmingly typified, powers of imaginative intuition’.

Moreover, according to Sacks (1984: 25), our imagination is limited by our perception of what we believe could permissibly be produced in talk-in-interaction:

‘however rich our imaginations are, if we use hypothetical, or hypothetical-typical versions of the world we are constrained by reference to what an audience, an audience of professionals, can accept as reasonable. That might not appear to be a terrible constraint until we come to look at the kinds of things that actually occur’.

A more important reason why CA and the Firthian approach repudiate the use of intuited or experimental speech data, however, is that both of these approaches are concerned with making analytic claims about language that are grounded in the details and relevancies of the interactional conduct of the participants-in-talk. This standpoint is evidenced in the following quotes, taken from scholars working within the Firthian and CA traditions respectively:

‘we must apprehend language events in their contexts as shaped by the creative acts of speaking persons’ (Firth 1951: 193).

‘the empirical conduct of speakers is treated as the central resource out of which analyses may develop’ (Heritage 1984a: 243).

1.4.5 Participant Orientation as a Basis for Theorising

In CA and the Firthian approach, the observable interactional behaviour and orientations of the parties-to-talk are used ‘as a basis for theorising’ and comprise the analytic base from which the categories of analysis are extrapolated and warranted (Sacks 1984: 25).

This basing of theory on observation has been pointed out by Firth and Schegloff (the latter being a founding scholar and leading figure in CA):

‘We can only arrive at some understanding of how it [language] works, if we establish with certainty that the facts of speech we are studying can be observed or regarded in actual patterns of behaviour. We must take our facts from speech sequences...operating in contexts of situation which are typical, recurrent and repeatedly observable’ (Firth 1935b: 35).

‘When the data are drawn from the materials of human interaction, and display the orientations to relevance which inform the conduct of the participants themselves, then the problems of the indefinite extendibility of description can be resolved ... by reference to the relevancies to which the *participants themselves* show themselves to be oriented’ (Schegloff 1988: 21).

One methodological consequence of this stance is that, as noted by Heritage (1989: 23), ‘findings will ultimately be answerable not merely to statistical treatments of aggregates of data, but also to the detailed explication of singular instances’ (see also Schegloff 1987a, 1993; Wootton 1989). In other words, if a participant can be seen to orient to a particular feature then that feature is regarded as being *significant* regardless of how many times it may occur in the interaction. As pointed out by Haakana (2002: 228): ‘For the interactants the question is not how ... [the item under investigation] is statistically distributed, but what happens here and now in the interaction’.

The empirical grounding of the analysis in accordance with the demonstrable conduct of the participants also places constraints on the allowable statements of the analyst. Consequently, both CA and the Firthian approach take it as axiomatic that preconceived assumptions and premature theory constructions about language should be avoided, as all analytic statements must be answerable to the data in hand (see Firth 1957: 199; for

CA see Heritage 1989 and Schegloff 1987b). Moreover, both approaches also warn against and eschew appeals to psychological phenomena as a means of explaining language, arguing that only the actual observable conduct of the interactants should comprise the foundation of theory construction.

1.4.6 Summary

This section has shown that CA and the Firthian approach share strikingly similar views on language and its context. These include the belief that:

- talk-in-interaction is social action;
- investigations of language should be based on naturally-occurring talk-in-interaction rather than contrived and idealised data;
- speech should be analysed as a contextualised production, as language meaning cannot be divorced from its context and vice versa;
- the interpretation of an action is shaped and renewed by its context of production;
- language is polysystemic;
- only the orientations of the interactants should be used as a basis for theorising;
- premature theory construction should be avoided.

The similarities in opinion between CA and the Firthian approach are fortunate, as although Firth bequeathed many great ideas about language and its analysis, he did not provide much exemplification or elaboration on how to implement these ideas. CA therefore provides the Firthian-oriented linguist with a rigorously empirical framework and a methodological toolkit with which many of Firth's ideas and beliefs about language can be incorporated, investigated and implemented.

One of the most fundamental similarities between CA and the Firthian approach is the belief that speech should be analysed as a contextualised production embedded in its natural habitat of social interaction. Both these approaches regard the *internal*, situated context of speech production as providing its interpretation for the parties-to-talk and, subsequently, for the analyst. In comparison, although the Variationist approach concerns itself with the examination of socially embedded speech productions (unlike the traditional generative approach), it typically appeals to contextual features *external* to the speech production. These different internal and external invocations of context in CA and the Variationist approach will be examined and compared in the next section (1.5).

1.5 Invocations of Context in CA and Variationist Studies

The Variationist invocation of context differs from the Firthian and CA invocations, as the Variationists typically appeal to *external* contextual features of the talk-in-interaction, such as the speaker's age or social class, whereas the Firthian and CA approaches appeal instead to *internal* contextual features of the sequential organisation of talk (see 1.3 above). This section will examine these two different invocations of context: external and internal, otherwise known as distal or proximate (1.5.1). After, a fragment of talk will be analysed in which the relevance and procedural consequentiality of different types of contextual features will be exemplified (1.5.2).

1.5.1 Distal vs. Proximate

As argued by Schegloff (1992: 195), a leading figure in CA, the term context has frequently been used in investigations of language to denote *two different types* of context: (1) the external or distal context and (2) the intra-interactional/discourse/proximate context. The former type of context (i.e. the distal

context) refers to a participant's *social structure* such as their age, sex and ethnicity. In contrast, the latter type of context (i.e. the proximate context) encapsulates:

‘the sort of occasion or genre of interaction which participants, by their conduct, make some episode be an instance of, the sorts of sequences of talk or courses of conduct in which particular events may occur (stories, request sequences, etc.), the capacity in which participants act relative to the episode in progress (e.g. as the initiator of a conversation or a topic, or its recipient), etc’ (Schegloff 1992: 195).

In other words, the proximate contextual factors of an utterance concern what the participants are actually *doing* in and through their interaction whereas the distal contextual factors concern the *social characteristics* of the parties-to-talk.

The macro-oriented, Variationist approach, outlined in 1.3, seeks to relate linguistic-phonetic patterns to *distal* contextual features of the speech production. Recall, for example, that the social characteristics of a speaker such as their age, gender and social class are typically invoked in Variationist studies in order to explain and account for patterns in language use. One potential difficulty of appealing to external contextual factors, however, is that, as noted by Schegloff (1992), the number of features which can be invoked is virtually infinite. For instance, factors such as the age, sex and class ‘composition of an interaction’ are just the tip of the iceberg with a plethora of others, such as the ‘ecological, regional, national and cultural settings’ of the interaction, also being potentially relevant (ibid: 195). This multitude of distal contextual features can therefore lead to difficulties when one wishes to work out ‘*which* of the possible characterizations of context ...can be shown to be *relevant*’, as there are so many with potential relevance (ibid: 195; emphasis in original). Moreover, even though some of the distal contextual features may be found to be relevant to the parties-to-talk, this does not necessarily entail that they are ‘procedurally consequential’; that is, that they have consequences for the ‘shape, form, trajectory, content or character of the interaction that

the parties conduct' (Schegloff 1991: 52-53). Thus, it is not enough to simply identify that a particular distal characterisation of context is relevant to the participants in the interaction. Instead, it must also be shown that this characterisation has a direct impact on how the talk sequentially unfolds.

Unlike the Variationist approach, CA considers the contexts which are relevant to the parties to talk as being the *only* relevant contexts (see Schegloff 1987b, 1991 for discussion). Schegloff (1992: 196; emphasis added) argues that: 'If one is concerned with understanding what something in interaction was for its participants, then we must establish what sense of context *was relevant to those participants*, and at the moment at which what we are trying to understand occurred'. In other words, Schegloff is claiming that rather than looking *outside* of the interaction by appealing to distal contextual features of the interaction, we must instead look inside at the internal, proximate contextual features and work out which of these mattered and were significant for the participants.

It is important to point out that the CA approach is *not* claiming that distal contextual features do not exist nor that they may not influence a course of social interaction. Rather it is instead arguing that the distal contextual features of an interaction are only deemed relevant to the analysis if the parties-to-talk display an orientation to them as being relevant as the interaction unfurls. In the case of the claims made by Variationists, then—that a person's linguistic properties are governed by social factors such as their age, gender and social class—a CA scholar would instead argue that evidence would need to be found in the talk itself that these distal features are oriented to by the participants in the interaction.

One study which invokes distal contextual features and is of particular interest to this thesis because of its claims about the use of clicks is that conducted by Benor (2004). In this work Benor argues that people in the social category {+American, +Orthodox,

+Jew} produce clicks in talk-in-interaction to index their ethnicity. However, from a CA perspective, evidence would need to be found in the unfurling sequential organisation of the talk that these distal contextual features are oriented to by the interactants as being relevant and consequential. That people in a different social category, e.g. {+English, -Orthodox, -Jew}, also regularly produce clicks, however, raises questions as to whether the clicks are indeed related to the distal contextual feature of ethnicity or whether they are instead part of the proximate contextual design of the talk. The research in this thesis demonstrates that clicks can indeed be related to the proximate contextual features of the interaction, as they can be employed to index the interactional structure of the talk in which they are situated.

To this point, the discussion of CA-oriented investigations has focused on the way they appeal only to the proximate contextual features of talk which have demonstrable relevance and procedural consequentiality for the participants in the interaction. This is not to say, however, that the distal contextual features can never be invoked. Only that if they are, then they must be shown to be relevant and consequential for the parties-to-talk. An episode of interaction in which both proximate and contextual features can be shown to have demonstrable interactional significance is discussed in the next section (1.5.2).

1.5.2 Evidencing Proximate and Distal Contextual Features

In this section, a fragment of talk is examined in order to exemplify the relevance and procedural consequentiality of some of the proximate and distal contextual features of a course of social interaction. In the first part of this discussion, an outline of some of the proximate contextual features of the interaction which are relevant and procedurally consequential in the *opening* sequence of the call will be provided (a). Next, the distal contextual features of the interaction will be outlined and it will be discussed that most of these features are typically not relevant or procedurally consequential for the same

opening stretch of talk (b). It will also be shown, however, that it is not always the case that the distal contextual features are not relevant or important to the way the interaction unfolds, as there is one distal factor in the fragment which is particularly relevant throughout most of the call.

a) The Proximate Contextual Features

Fragment 1 is taken from a telephone-call produced on an American radio phone-in show called The Giant Sixty Eight KNBR, which people ring to share their views on particular issues (see Appendix 1 for transcription conventions).⁸ The issue being discussed in this call is the 1991 Gulf War, which was in progress at the time of the show's airing. There are two participants in the call: the caller (Don) and the presenter (Pre). Don, the caller, has rung the radio station to share his views on the war. The episode of interaction begins with the presenter introducing Don, (Don on the line from San Francisco (L1)) and stating that he is now on air (you're on the giant sixty eight KNBR (L2)). In response, Don orients to the presenter's turn as (a) indicating that it is his (Don's) turn to speak on the show and (b) as being a greeting, as evidenced by his subsequent greeting and other-identification: hi Leo (L3). Moreover, Don's turn (L3) also functions to index that he is in fact the named person from Line 1 (and that there is not a problem on the phone-line). What Don does *not* do, then, is: (1) remain silent (after the presenter has finished his introductory sequence) or (2) go straight into his reason-for-call without producing any prefatory material.

After Don's greeting, a microgap occurs (L4) in which neither of the participants produce any talk. Then, both Don and the presenter take up the next turn simultaneously, Don with the particle *uh* (L5) and the presenter with the lexical item *thanks* (L6). Don's production of *uh* suggests that he is going to produce further talk

⁸ This show is based in Berkeley, California (Couper-Kuhlen 1998: 3). Thanks to Elizabeth Couper-Kuhlen for making this data available and to Richard Ogden for bringing this fragment to my attention.

(Sacks 1992a: 497; Jefferson 1984a: 216) however at this point in the call he curbs his turn to allow the presenter to continue out of overlap (L6). The presenter continues his turn stating *thanks for calling Don* (L6), the first two words of which are produced in overlap with Don's particle and the last two of which are produced in the clear. From its lexical makeup, the presenter's turn in Line 6 appears to function as an appreciation to Don for having called the show: *thanks for calling Don* (6). However, in this location Don does not orient to this turn as being an appreciation. Instead, he treats it as providing him with a license to subsequently produce his opinions about the war and thus present his reason-for-call. This is evidenced in Don's immediately subsequent talk (L7-20) in which he produces an extended multi-unit turn, which concerns a telling of his views on the then current war in Iraq. Note that this extended sequence is initiated with Don's informing: *I am a retired air force officer* (L7).

It is worth pointing out once again that Don orients to the presenter's 'appreciative' turn in Line 6 as providing him with a go-ahead to produce his reason-for-call because of its sequential positioning. In other words, Don's interpretation of the presenter's turn is achieved from its embedded context of production rather than simply its lexical format. This context-sensitive interpretation of language can be seen more clearly by examining the sequential relevance and procedural consequentiality of an almost identical phrase produced by the presenter in the closing section of the call. In the first part of the closing section the presenter provides an assessment to support a preceding assessment by Don (*well if you're gonna fight this is the way to do it* (L61)), which is subsequently agreed with by Don in third turn position (*I agree* (L62)). Then, the presenter states *thank you Don* (L63), a stretch which has a similar lexical composition to that produced by the presenter in the opening sequence (*thanks for calling Don* (L6)). Note, however, that unlike Don's response to the presenter in the opening section of the call, in the closing section Don treats the presenter's turn in Line 63 as doing (1) a closing down rather than an opening up of the call and (2) an

appreciation. This is evidenced by Don's non-topical response of okay thank you (L64), which serves to both accept the presenter's offer of closure and also reciprocate the appreciation.

Fragment 1: /10gulfwar/

01: Pre: Don on the line from San Francisco
02: you're on the giant sixty eight KNBR
03: Don: .h hi Leo
04: (.)
05: Don: [u h : :]
06: Pre: [thanks for] calling Don
07: → Don: **I am a retired air force officer**
08: and I just wanted to: uh put out the word that
09: basically I appreciate the way the (.) president is
10: → handling this thing **he's going all out no holds barred**
11: * *I've been in Vietnam I've been in other things .hh*
12: → **and when you go in there with your hands tied**
13: → **behind your back-** or you get congress involved
14: it's it's (.) kind of scary because you really (.)
15: aren't in control of the situation
16: a:n right now (all) congress will stay out of it
17: → they've turned it over to the military and **let them**
18: → **(.) you know take the action**
19: → **go in there and clean house (.)**
20: I think we'll be alright
21: Pre: you've seen air action before have you
22: ever seen anything like what happened tonight
23: * Don: *no I tell you what I've seen uh raids in north*
24: * *Vietnam and things like that and I don't believe*
25: that uh I've ever heard of anything close to this
(15 lines omitted)
41: Pre: were you surprised at the effectiveness of the technology
42: it much of it unproven
43: Don: *no I really wasn't I I've been around (all) like*
44: * *say I was in the military over twenty years and I I*
45: (.) a lot of our weapon systems yes had not been
46: proven (.) but they've been tested
(15 lines removed)
61: Pre: well if you're gonna fight this is the way to do it
62: Don: I I agree
63: Pre: thank you Don
64: Don: okay thank you
65: Pre: appreciate the call

b) The Distal Contextual Features

Some of the distal contextual features which could potentially be invoked in this call include gender, age, nationality, social class and ethnicity. For example, both of the participants could arguably be placed in the contextual categories of: {man},

{middle/late aged}, {American}, {middle/upper-middle class}.⁹ These distal categories, however, are not, in the first instance, oriented to by the interactants as being either relevant or procedurally consequential for the way the interaction unfolds. For example, Don's interpretation of the presenter's 'appreciative' turns in L6 and L63 is not determined by the presenter's social attributes but rather is achieved from the sequential placement of the turns in the interactional structure of the talk. Thus, as pointed out by Schegloff (1987b: 217), by seeking to ascribe external social attributes of the parties-to-talk to the organisation of talk-in-interaction we run the risk of underspecifying the interactional phenomena. This is not to say that one should *not* include these macro-social attributes in micro-analyses of talk-in-interaction - only that their inclusion must be validated by the participants' orientations to them as being relevant. As argued by Schegloff (ibid: 216):

'what is needed is the capacity to specify technically the parameters of the relevant organization of action or interaction through which macroattributes have whatever different effects they have, if any'.

One distal contextual feature which can be shown to bear relevancies and procedural consequentialities for much of the call in Fragment 1—and can be warranted by the participants' orientations to it—is Don's membership in the social group of {+retired air force officers}. Recall that Don states his membership in this distal category as a preface to his reason-for-call: I am a retired air force officer (L7). His statement in this initial location serves to make this social attribute immediately relevant, as it indexes that Don himself regards this feature of himself as being relevant for the purposes of the call.¹⁰ It is not only relevant in this position, however, but affects most of how the interaction unfolds for both Don and the presenter. Observe, for

⁹ These are only rough estimations of the age and social class of the interactants for the purposes of illustration. The author recognises that far more detailed investigation would be needed if one were to invoke these categories in an analysis.

¹⁰ Note also that this radio show occurred on the day that the *air* campaigns began in the 1991 Gulf War thus adding extra relevance to Don's statement that he is a retired air force officer.

example, that it affects the terminology that Don uses (see e.g. the following phrases, which are emboldened and initiated in the transcript with an arrow (as is Don's preface in L7): **he's going all out (L10)**; **no holds barred (L10)**; **when you go in there with your hands tied behind your back (L12-13)**; **let them take the action (L17-18)**; **go in there and clean house (L19)**). Moreover, Don also makes reference to his membership in the distal category *ex-air force officer* at three other points in the call (all of which are italicised and initiated with a star): *I've been in Vietnam I've been in other things (L11)*; *I tell you what I've seen raids in north Vietnam and things like that (L23-24)*; *I was in the military over twenty years (L44)*. That Don re-references his membership in this social group at three additional and separate points in the call further evidences the proposal that he regards this distal feature of himself as being relevant to the interaction.

Importantly, the distal contextual feature of Don being a retired air force officer is not only relevant and procedurally consequential for Don, but also for the presenter. Notice that after Don's initial, long stretch of talk (L7-20), the first contribution from the presenter is the question: *you've seen air action before have you ever seen anything like what happened tonight (L21-22)*; which is soon followed by: *were you surprised at the effectiveness of the technology it much of it unproven (L41-42)*. These questions, each of which are underlined in the fragment, evidence that the presenter treats Don's characterisation of himself as being a former air force officer as providing him (Don) with specialist knowledge with which to answer these questions and that it is therefore a relevant and procedurally consequential distal contextual feature of the interaction.

It is important to note that although Don's membership in the external category of retired air force officer is relevant to the way in which most of the call unfurls, its relevance is also variable. In the opening stretch (L1-6), for example, this distal feature

is not a relevant property of the interaction, as the parties-to-talk do not orient to it at this point in the call. Similarly, it is also not relevant in the final closing section from L63-65 in which the interactants work together to collaboratively close down the call. Thus, just because a particular category is relevant at some point in talk does not mean that it is relevant throughout the entirety of the interaction. Rather, categories in talk-in-interaction can have variable relevance at different places in structure (see Local 2003).

1.5.3 Summary

This section has outlined the different contexts invoked in the CA approach compared with those invoked in Variationist studies. The Firthian and CA approaches routinely appeal to *internal* contextual factors when explaining patterns of language use, otherwise known as proximate features, whereas the Variationist approach typically invokes contextual factors *external* to the speech production, commonly referred to as distal contextual features. The (non-) relevance and procedural consequentiality of these different types of contextual feature have been exemplified using a data fragment.

Now that the Firthian (1.2) and Variationist (1.3) approaches have been outlined, a comparison has been drawn between the CA and Firthian school of thought (1.4) and the different contexts invoked in Variationist and CA studies have been exemplified (1.5), it is possible to highlight how these different approaches have influenced this research and to present the directions and motivations of the study.

1.6 Directions of and Motivations for Research

This thesis focuses upon the interface between the linguistic-phonetic and the interactional-sequential design of naturally-occurring English talk-in-interaction. In particular, it examines the relationship between the distribution of clicks and the interactional organisation of the talk in which they are produced. Until now, clicks have

typically been regarded as functioning only paralinguistically in English, indexing various emotional and attitudinal states of the speaker; there is therefore no treatment of them in English phonology. The suggested states that clicks are said to convey include disapproval (Ladefoged 1982: 124; Crystal 1987: 126), annoyance (Abercrombie 1967: 31; Ball 1989: 10), irritation (Gimson 1970: 34), exasperation (Laver 1994: 175), impatience (Laver 1994: 175), regret (Clark & Yallop 1990: 18), sympathy (Gimson 1970: 34) and encouragement (Gimson 1970: 34; Abercrombie 1967: 31; Ladefoged 1982: 124; Laver 1994: 177). Another paralinguistic use of clicks is that of providing a kiss, such as ‘the gentle, pursed-lips types of kiss that one might drop on one’s grandmother’s cheek’ (Ladefoged 1982: 124) or a ‘kiss at a distance’ (Crystal 1987: 126). The phonetic properties of these paralinguistic functions of clicks in English are said to differ according to their function. For example alveolar or dental clicks are said to be used express irritation, sympathy (Gimson 1970: 34), regret and disapproval (Clark & Yallop 1990: 18; Ladefoged 1982: 125, 2001: 155), impatience and exasperation (Laver 1994: 175) and lateral clicks to encourage animals, particularly horses (Gimson 1970: 34; Crystal 1987: 126; Clark & Yallop 1990: 18; Laver 1994: 177).

Although various works have proposed many different paralinguistic functions of clicks, these works share one important commonality: they are not based on naturally-occurring conversation. There are, however, two notable exceptions in which the functions of clicks are examined in actual conversation. These exceptions are the works by Ward (forthcoming) and Benor (2004). In attempting to provide a sound-meaning mapping for non-lexical sounds in American English, Ward (forthcoming: 22) investigated clicks in American-English conversation and claimed that: ‘The meaning of tongue clicks can be subsumed under the term personal dissatisfaction’; Benor (2004: 189) argued that ‘hesitation clicks’ are often produced in conversation by Orthodox Jewish American-English to index their ethnicity. However, although these works both make claims about the functions of clicks based on naturally-occurring English talk-in-

interaction (and take seriously the important roles that non-lexical items may undertake), they both still regard clicks as functioning only paralinguistically. Moreover, their claims are based on intuited categories, rather than being warranted by the observable conduct of the participants.

With all of the previous claims about the functions of clicks being based on anecdotal observations or the analyst's own intuitions, many of the claims are not borne out when one examines the occurrence of clicks in naturally-occurring talk-in-interaction. This is because some of the clicks cannot be found to undertake any of the proposed paralinguistic functions outlined above. A case in point is provided in line 3 of fragment 2. Observe in line 3 that Lesley produces a bilabial click (transcribed phonetically as [⊙]) between *yep* and *okay then*. In this location, the click cannot be said to undertake any of the proposed paralinguistic functions, as there is no evidence in the surrounding talk that the speaker is, for example, irritated, annoyed or exasperated. Instead, the click can be seen to mark a structural location in which sequence transition occurs. This is evidenced by its embedded context of production, as the sequence which precedes the click (up to and including *yep*) concerns a telling of someone's attendance at an upcoming meeting (L1-3); whereas subsequent to the click, there is an immediate sequence shift into the closing section of the call (L3-8) (which is initiated with *okay then* (L3)). Thus, rather than being engaged in the signalling of an emotive state, this click appears, instead, to be engaged in the phonetic design of sequence transition.

Fragment 2: Holt.C.1985.6/splendid/

01: Les: so she's comin::g to the f:irst one after Christmas
 02: Bod: Oh: splendid
 03: → Les: **yep (0.2) [⊙] okay the[n**
 04: Bod: [righ[t well I shall]see you
 05: Les: [see you later]
 06: (.)
 07: Les: B[ye bye:
 08: Bod: [Bye

In addition to difficulties in mapping the proposed paralinguistic functions of clicks to clicks produced in talk-in-interaction, problems can also be encountered when one wishes to *distinguish* between the different paralinguistic functions of clicks, as many of them, such as irritation, annoyance and exasperation, are strikingly similar. This raises questions regarding whether the various proposed types of click are actually different from each other (or whether it is simply the terminology used or the analysts' intuitions which differ). For example, are the clicks which are used to show 'annoyance' different from and used in different places to the clicks which are used to show 'impatience'? Additionally, as noted by Crystal (1969), how can the attribution of clicks to various emotional and attitudinal (i.e. paralinguistic) states of the speaker be constrained?

One way in which all of these difficulties experienced with paralinguistic accounts of clicks can be overcome is by the adoption of a rigorous methodology in which (1) the data are drawn from naturally-occurring talk-in-interaction and (2) any claims made about the function of clicks are grounded in—and constrained by—the observable orientations of the participants. The methodology of Conversation Analysis therefore provides an appropriate methodology with which an empirical analysis of clicks in talk-in-interaction can be undertaken. It has therefore been adopted in this research and its methods of investigation have proved fruitful in the identification of mappings at congruent levels between the occurrence of clicks and the syntagmatic structure of English talk-in-interaction. To the best of the author's knowledge, this thesis presents the first systematic, empirical study of clicks in English talk-in-interaction, in which an interrelationship is demonstrated between the distribution of clicks and the interactional organisation of the talk in which they are produced.

The motivation for examining the phonetics-interaction interface in this thesis stems from the fact that, as outlined in the introduction, our knowledge of the linguistic-phonetic organisation of naturally-occurring talk-in-interaction is still relatively limited.

Even more limited is our understanding of the ways in which parties-to-talk design and manipulate particular constellations of phonetic parameters for interactional purposes. These limitations in knowledge result from talk-in-interaction being understudied in linguistic research, as was noted by Firth in 1935:

‘Neither linguists nor psychologists have begun the study of conversation, but it is here we shall find the key to a better understanding of what language really is and how it works’ (1935a: 32).

In 1946, Bolinger also highlighted a lack of linguistic research into details of talk-in-interaction, when he argued that there has been:

‘a close concentration upon a limited number of behaviour patterns, the latter suggestively reminiscent, in their selection for ease of recordability, of the ‘written forms’ from which we were supposed to have been emancipated. It is only by a return once more to the whole of communicative behaviour, with energies of linguists more evenly distributed, that we shall avoid the over-growth and premature refinement of one or two component parts’ (p. 92).

Similar views were echoed by Abercrombie in 1965 when he highlighted a deficiency of research on conversation as a result of linguists concentrating on the study of ‘spoken prose’ (p.4). Up to the present day, however, the linguistic-phonetic properties of naturally-occurring talk-in-interaction are little studied compared to the plethora of research conducted on idealised speech data. The Firthian/CA and Variationist approaches outlined above (see 1.2 to 1.4 respectively) are therefore notable exceptions, as each highlight and stress the importance of examining actual, socially situated speech productions (however the main concern of the latter approach in examining these productions is the phonetic encoding of sociolinguistic variation whereas in the former, it is the encoding of interactional functions/meanings).

Additional notable exceptions which do indeed typically investigate naturally-occurring talk include much of the discourse analytic research (see e.g. Schiffrin 1987; Schourup 1985; Fox Tree & Schrock 2002) and studies of English intonation, particularly that oriented to the intonation, or more specifically the pitch, of English discourse (see e.g. Crystal 1969; Crystal & Davy 1975; Cruttenden 1986; Pierrehumbert & Hirschberg 1990; Tench 1996). However, although these fields of research routinely examine the properties of informal conversations rather than relying on analyst-driven intuitions they are subject to a number of shortcomings which can be attributed to the following factors: the analyst's reliance on intuition rather than participant orientations when establishing the functional categories of analysis; the focus on one pre-specified feature such as a certain 'discourse marker' or, in the case of the intonation studies, pitch at the expense of other components or phonetic parameters; and the attribution of monosystemic statements in which one overall function is sought to describe the meaning of the discourse marker/intonation contour under investigation (see Local et al. 1986: 411-13, for further discussion on some of these and other limitations of 'discourse intonation' studies).

The direction of the research in this thesis is greatly influenced by the ideas of Firth and his co-workers and as such, regards itself as a modern-day Firthian approach to the study of talk-in-interaction. It aligns, for example, with the Firthian insistence that language should be analysed as a contextualised production embedded in its natural habitat of social interaction. Moreover, it also considers the context in which language is produced as being inseparable from its meaning and its meaning as being inseparable from its context. It therefore appeals only to the context embedded within the internal structuring of the talk rather than to externally invoked contexts when explaining patterns of language use, the latter of which is commonplace in and indeed the endeavour of Variationist studies.

This thesis is also massively influenced by the CA approach, as it holds many strikingly similar views on language to the Firthian approach (see 1.4 for a detailed comparison). For instance, as with the Firthian view, the CA approach maintains that investigations of language should be based on naturally-occurring talk-in-interaction and that meaning in talk is contextually shaped and determined. Unlike the Firthian tradition, however, which paid much lip-service to the investigation of talk-in-interaction but instead focused mostly on other material, CA researchers have focused entirely on documenting the orderly details of talk-in-interaction. To this end, CA scholars have devised a rigorously empirical, methodological framework within which the details and organisation of talk-in-interaction can be explored. The strict empirical stance embodied within the CA framework has resulted in this tradition obtaining one of its original and most fundamental goals; that of achieving a ‘a naturalistic observational discipline that could deal with the details of social action(s) rigorously, empirically and formally’ (Schegloff & Sacks 1973: 289-290). It has therefore successfully overcome the trap of which Firth warned, of developing nothing more than ‘a loose linguistic sociology without formal accuracy’ (Firth 1935a: 31, 1fn).

CA studies have shown that detailed, sequential investigations of naturally-occurring talk-in-interaction can uncover a plenitude of interactive practices which participants systematically employ and orient to in talk-in-interaction. Linguistic-phonetic research which has employed CA techniques of analysis to examine the phonetics of talk-in-interaction, and has hence responded to the CA calling for linguists to investigate the linguistic exponents of interactive practices (Sacks et al. 1974), has fruitfully excavated and explicated the interactional implications and sequential organisations of an array of linguistic-phonetic events (see e.g. French & Local 1983; Local & Kelly 1986; Local et al. 1986; Local 1996, 2004; Couper-Kuhlen 1993, 1996, 2004b; Tarplee 1993; Wells & Peppé 1996; Ogden 2001, 2003, Wright 2001; Local & Walker 2005; Curl 2002; Curl et al. 2004; Walker 2004a, 2004b; Ogden et al. 2004). This approach, which combines

phonetics and CA, was founded by John Local and Peter French in the 1980s (French & Local 1983). Since its inception, a growing number of linguists have adopted its methodology and have demonstrated that by: (i) adopting an interactive approach in which the categories of analysis are empirically warranted by the interactional conduct of the parties-to-talk; (ii) employing a parametric phonetic analysis in which no details are excluded as being irrelevant until they are shown to be irrelevant for the participants; and (iii) regarding language as being polysystemic, fruitful advances can be made in our understanding of the linguistic-phonetic organisation of talk-in-interaction. This study contributes to this expanding and exciting body of interactionally-oriented linguistic-phonetic research. Its focus on clicks provides an excellent case study of the CA notion that no order of detail should be dismissed a priori as disorderly or unimportant.

On a final, introductory note it is important to point out that although the combination of phonetic and sequential-interactional (CA) techniques is a relatively new and highly radical linguistic approach, the concept of detailing the linguistic-phonetic exponents of particular interactional events in social interaction is in fact an idea which can be traced back to Firth, as evidenced in the following quote:

‘I want to make it clear that the linguistic systems and structures are related to the systems and structures in the events, relevant objects and people and what they are doing’ (Firth 1956b: 91).

However, as noted in 1.2, Firth did not provide much detail regarding how one can identify the interactional systems and structures of talk hence practically no advances were made in our understanding of the phonetics-interaction interface until, inspired by Firthian concepts, John Local, Peter French and co-workers employed the empirical notions of CA to devise their interactionally grounded approach to linguistic-phonetic analysis. The CA approach of delimiting the interactive categories according to the orientations of the interactants therefore presents the Firthian linguist with a particularly

sharp tool with which the systems and structures of language can be unearthed. An additional notion which percolates throughout the interactive phonetic approach and which can be traced back to the Firthians is that of examining the interface between the phonetic and the interactional structure of contextually embedded talk-in-interaction:

‘...it is true that it [linguistics] is peculiar amongst sciences in standing astride two streams of phenomena – on one side the phonic material which constitutes speech, and on the other the practical situations in which speech operates. *These situations or ‘contexts’ may be considered as functions of the phonic material which operates within them, or in other words as the ‘meanings’ of that material...* It is only when we analyse phonic material by reference to its contextual function that those peculiarly systematic statements become possible which are characteristic of linguistics. An analysis which ignored such function, and concerned itself solely with the phonic material, might, for example, define its units in terms of arbitrary vocal intervals and the pauses between them; for linguistics, as Professor Bazell has observed, ‘the unit so defined would have the interesting property that nothing what-ever profitable could be said about it’ (Allen 1957: 23; emphasis added).

In this quote Allen is arguing that the levels of phonetics and interaction should always be analysed hand-in-hand otherwise we run the risk of having improperly defined units of analysis which tell us nothing about the organisation of the phonetics. The belief that the interactional and phonetic details of talk-in-interaction must be examined in tandem is one which is fundamental to and permeates throughout the research in this thesis.

1.7 Outline of Thesis

The remainder of this thesis is organised as follows:

Chapter 2 details the data and methodologies which are employed in the three studies presented in chapters 3, 4, and 5. The data are taken from naturally-occurring telephone

talk-in-interaction in which the interaction had real interactional significances for the participants. The methodology combines the sequential techniques of CA with impressionistic observations and instrumental analyses.

Chapter 3 presents an analysis of multi-unit first closing turns, a device which speakers regularly employ to shift the direction of their talk-in-interaction from some on-topic talk into the closing section of the call. These turns are shown to have systematic and recurrent regularities in their sequential, interactional and phonetic designs. One striking feature of these turns is the routine occurrence of an audible click in the boundary between the two units/sequences of talk.

Chapter 4 explores the occurrence of clicks in sequence boundaries in talk-in-interaction in more detail. It demonstrates that clicks regularly occur in the boundary between the closing down of one sequence and the initiation of a following disjunctive sequence. In this location, these clicks are shown to demarcate the onset of a new sequence from the prior and are thus termed *new-sequence indexing clicks*. Sequential, interactional and phonetic evidence are provided to support this claim.

Chapter 5 examines another sequential and interactional environment in which clicks are regularly produced: word searches. The environments of these *word search clicks* are shown to have markedly different sequential, interactional and phonetic designs from the environments of the new-sequence indexing clicks. These design features provide evidence of the action of word searching.

Chapter 6 presents a detailed summary of the thesis, outlining the work undertaken in each chapter, the key findings and the findings specific to clicks. It discusses some of the limitations of the study which, in turn, are shown to influence the proposed directions for further research. Finally, the implications of the study are presented.

The findings of this research make a significant contribution to our understanding of the phonetic and interactional organisation of everyday social intercourse. They highlight the fruitfulness of implementing context-bound phonetic investigations alongside interactional analyses, since previously unobserved orderly and recurrent patterns in the phonetics-interaction interface can be clearly identified.

CHAPTER TWO

DATA AND METHODOLOGY

2.1 Introduction

This chapter details the data and methodologies employed in the analysis in chapters 3 to 5. First, the data are presented (2.2), followed by the corpora (2.2.1) and the transcription conventions (2.2.2). Next, the approach of Conversation Analysis (CA) is discussed (2.3), as its methodology underpins the analysis conducted in this thesis - details are provided about the use of participant orientations when (a) constructing and evidencing arguments about the design of social interaction (2.3.1) and (b) when measuring the significance of an observation in CA (2.3.2). Finally the phonetic methodology is outlined (2.4), which comprises both impressionistic (2.4.1) and instrumental analysis techniques (2.4.2), and details are given regarding the phonetic investigations undertaken and their presentations in the analysis chapters.

2.2 Data

This section presents the corpora used (2.2.1) and the transcription conventions employed (2.2.2) in the study.

2.2.1 The Corpora

The data are taken primarily from six corpora: the Holt corpus, the Heritage corpus, the Rahman corpus, the NB corpus, the SBL corpus, and the Pline corpus. Details of these corpora are given in Appendix 2. These corpora were recorded between 1960 and 2001

and comprise around 18 hours of naturally-occurring telephone talk-in-interaction. The calls are mostly produced between friends and/or family members and typically contain only two interactants at any one time. Four of the corpora—the Holt, Heritage, Rahman and Pline corpora—are composed entirely of British-English speakers, and two—the NB and SBL corpora—contain only American English speakers.¹

It is important to point out that the corpora used comprise numerous speakers with a variety of different social characteristics. These social differences include gender, age, ethnic/national background and social class. The interactants also represent a diversity of accents: some of these differences are national (e.g. English-English (UK) in the Holt corpus; Welsh-English (UK) in the Rahman corpus; American-English (USA) in the NB and SBL corpora) while others are regional, such as the non-standard English typically spoken in Somerset (England (UK) in the Holt corpus) and West Yorkshire (England (UK) in the Pline corpus). None of these different social, regional and national characteristics of the speakers, however, were found to influence the findings in this thesis, as the patterns observed hold across a vast and diverse set of speakers, regardless of their social and national makeup. For example, a young British-English male and an older American-English female were each found to produce clicks when searching for a word. This suggests that the results in this research are not specific to any given social or regional English language community but rather hold more generally across a large number of speakers with diverse social and regional attributes.

2.2.2 Transcription Conventions

All the data were transcribed using the transcription conventions given in Appendix 1, which were devised by Gail Jefferson in the 1960s. Five of the corpora (the Holt, Heritage, Rahman, NB and SBL corpora) had been previously transcribed in full using

¹ There is one speaker in the Holt corpus who does not appear to be a native English speaker.

the conventions given in Appendix 1.² The Pline corpus had not been previously transcribed, as it was collected by the author – it was transcribed by the author in accordance with the documented conventions outlined in Appendix 1.

In this thesis, the transcriptions of the data fragments are relatively plain, as many of the original CA conventions for transcribing talk-in-interaction were not used. For example, standard English orthography was used rather than an altered quasi-analytic orthography, which tries to capture some of the phonetic properties of the speech production. Where fragments from the pre-transcribed corpora are presented, they are modified to include only standard orthographic representations of the speech. Utterances transcribed, for example, as *is 'ee Oy:rish* are therefore changed to *is he Irish*. In addition, conventions which are used in CA in an attempt to capture various prosodic features of talk, such as ‘a rising inflection’ (shown as {?}), ‘a continuing intonation’ (shown as {,}) and ‘an animated tone’ (shown as {!}), are also not included (Atkinson & Heritage 1984: xi). The result is a relatively simple transcription which indexes little more than the sequential ordering of the various components of the interaction.

The use of plain rather than detailed transcriptions, the latter of which contain a multitude of extra quasi-analytic information, is indicative of the methodological stance adopted - that one should approach the data with as few preconceived assumptions as possible. Moreover, it also reflects the author’s opinion that general orthographic transcriptions are a useful means of capturing the overall content and flow of the talk but should not be used either as an analytic tool in and of themselves and/or as a substitute for listening to the details of the actual talk-in-interaction. Components of the interaction which are found to bear relevancies and consequentialities for the interactants are thus subjected to detailed investigations in their own terms rather than

² Thanks are due to Gail Jefferson for transcribing the Holt and Heritage corpora.

being applied to a general, ‘all-inclusive’ transcription. This means that features such as, for example, falling pitch over a particular stretch of talk are not shown on the transcription but are instead examined and presented separately using the phonetic analysis techniques outlined in 2.4. This allows for phonetic aspects not adequately captured using the CA notation of pitch, such as pitch span and placement of pitch in a speaker’s pitch range, to be detailed (see Tarplee 1993: 82-87 and Walker 2004a: 38-44 for a discussion of some of the shortcomings of CA notation).

2.3 Conversation Analysis

The research in this thesis is novel, as it combines interactional and phonetic approaches to examine the phonetics-interaction interface of naturally-occurring talk-in-interaction. The methodology which comprises the interactional approach originates in Conversation Analysis (CA) (see 1.4 for a discussion of some of the principles of CA). CA is a rigorously empirical approach to the study of talk-in-interaction which evolved through the pioneering work of Harvey Sacks and colleagues, such as Emanuel Schegloff and Gail Jefferson, in the 1960s. Sacks’ main aim was ‘to see how finely the details of actual, naturally-occurring conversation can be subjected to analysis which will yield the technology of conversation’ (Sacks 1992b: 339). In attempting to yield this technology, Conversation Analysts have devised and employed the following four key assumptions (taken from Heritage 1989: 22):

- talk-in-interaction is structurally organised;
- contributions to interaction are both context-shaped and context-renewing;
- no order of detail in conversational interaction can be dismissed a priori as disorderly, accidental or interactionally irrelevant;
- analysis should be based on naturally-occurring talk-in-interaction.

These assumptions will not be further discussed here, as introductory material on CA is plentiful. The reader is referred to the following references for further information:

Levinson 1983; Heritage 1984a, 1985, 1989; Wootton 1989; Schegloff 1991; Psathas 1995; Hutchby & Wooffitt 1998; ten Have 1999; Drew 2005.

In order to excavate, explicate and empirically evidence the orderly details of the technology of talk-in-interaction, through which and by which interactants organise and sequentially manage their social interactions, CA scholars undertake detailed, context-bound, sequential analyses which proceed on a turn-by-turn basis as the components of the interaction emerge in real time for the participants (Heritage & Atkinson 1984). These researchers are concerned with detailing ‘the interactional accomplishment of particular social activities’ (Drew & Heritage 1992: 17). Five domains of evidence are, as noted by Wootton (1989: 244), typically appealed to in CA investigations:

- the relationship of the device to just prior turns;
- co-occurring evidence within a turn;
- discriminability of the interactional device;
- subsequent treatment of the interactional device in question;
- deviant cases in the use of the device.

Common to each of these types of evidence is that they are empirically grounded in the data and warranted by the actual interactional conduct of the participants. In fact, this interactional warranting of claims is one of *the* most fundamental aspects of CA, as the strict empirical stance of CA maintains that the categories of analysis must be data-driven and that any arguments constructed about the organisation of naturally-occurring talk-in-interaction must be evidenced in accordance with—and only with—the observable orientations of the participants. As argued by Wootton (1989: 247) with regard to the evidential domain of ‘subsequent treatment of the device in question’:

‘in order to make claims as to the job performed by a turn it is incumbent on the analyst to demonstrate that the person occupying next turn reveals, through the design of his/her action, an analysis of the prior turn which is consistent with that being proposed by the analyst’.

Such an approach therefore places strict constraints on the researcher, as only the observable orientations of the interactants are used, and allowed, ‘as a basis for theorising’ (see 1.4.5 above for more discussion). Research which has adopted CA’s strict methodological stance has identified a plethora of devices which interactants systematically employ and orient to in order to manage their interactions in social intercourse (see e.g. Schegloff & Sacks 1973; Sacks et al. 1974; Sacks 1992a, 1992b; Schegloff 1979b, 1986a, 1986b, 1987c; Jefferson 1974, 1979, 1980a, 1984b, 1988; Jefferson et al. 1987; Button 1990, 1991; Atkinson & Heritage (eds.) 1984; Button & Lee (eds.) 1987; Drew 1997; Drew & Holt 1995, 1998; Heritage & Raymond ms; Heritage 1998, 2002).

In the following section (2.3.1), data fragments will be used to exemplify the import of participant orientations, particularly their ‘subsequent treatment of the interactional device in question’ (Wootton 1989: 247), when constructing and evidencing arguments in CA. This will be followed by an outline of the use of participant orientations as a measure of significance in CA studies (2.3.2).

2.3.1 Participant Orientations as Evidence

Fragments 1 to 3 below exemplify how the sequential analysis techniques of CA (see Heritage & Atkinson 1984) and its ‘participant-oriented proof procedures’ (Couper-Kuhlen 2001: 2) work in practice. Specifically, these fragments illustrate that the lexical item *what* can undertake a variety of functions according to its sequential placement, as evidenced by the participants’ subsequent treatment of the functionality of this ‘identical’ linguistic object (Wootton 1989). This discussion therefore also evidences a

motivation for the contextually-embedded, polysystemic approach adopted in this thesis, as the meaning/function of the word *what* is shown to be contextually determined.

Observe in fragment 1 that Marcia produces the word *what* (L7) in response to Donny's immediately prior pre-announcement (*guess what* (L6)) (Terasaki 1976; cited in Schegloff 1995a: 195; see also Terasaki 2004). In this location, Marcia's *what* displays her orientation to Donny's preceding turn as having news to tell and it provides a go-ahead to Donny to continue, thus also indexing that she (Marcia) has (potentially) not heard the news projected by *guess what*. Subsequently, Donny orients to Marcia's *what* as being a relevant response to his pre-announcement and one which indicates that he should, or can, continue with his telling. This is evidenced in Donny's next turn in which he delivers the projected news: *my car is stalled* (L8).

Fragment 1: taken from Schegloff (1995a: 193)

```
01:  Mar:  hello
02:  Don:  hello Marcia=
03:  Mar:  yea[:h
04:  Don:  [= ('t's) Donny
05:  Mar:  hi Donny
06:  Don:  guess what .hh
07:  → Mar:  what
08:  Don:  hh my ca:r is sta::lled
09:      (0.2)
10:  Don:  (and) I'm up here in the Glen
11:  Mar:  oh::
```

Unlike the orientations of the participants to *what* in the above fragment, observe their quite different orientations to the functionality of the 'same' lexical item in the following two fragments. In fragment 2, A's production of *what* (L8) after B's preceding question (*do you get to see Robin much* (L6)) engenders a repetition by B of B's preceding talk (see L10). This indicates that in this sequential context, B orients to A's *what* as being an open-class repair initiator (Drew 1997) which signals some kind of problem with the prior talk (see Curl 2002: 116 for a discussion of the interactional details of this fragment).

Fragment 2: taken from Curl (2002: 116)

01: A: if you want I can send this to you
02: (0.7)
03: A: I have a copy of it
04: (0.6)
05: A: [I- u-]
06: B: [hɪnʃu] do you get to see Robin much
07: (0.4)
08: →A: what
09: (0.2)
10: B: do you get to see Robin much
11: A: yeah once in a while cause she lives pretty near me

Contrastingly, in fragment 3, the child's production of *what* (L3) after her mum's preceding request (Mhairi will you take these things (L1)), does not receive a subsequent repair in the form of a repetition from the mother. Instead, in her next turn, the mother hurries the child with *come on quick* (L5), after which the child subsequently provides her reason for having produced the (preceding) repair initiator: *you're not saying please* (L7). This evidences that the child's production of *what* did not display a problem of hearing but rather was a claim to have not heard because of the lack of an appropriate form of politeness (see Drew 1997: 95 for further discussion on this point).

Fragment 3: taken from Drew (1997: 95)

01: Mum: Mhairi will you take the:se things
02: (0.6)
03: →Chi: wha:t
04: (1.8)
05: Mum: Come on quick
06: (.)
07: Chi: you're not saying plea::se=
08: Mum: =(no::(.) just take one)

It can be seen in the above data fragments, that the meaning/function of *what* is contextually embedded and sequentially determined - it emerges as the interaction unfurls between the participants and is evidenced by the way in which it is subsequently treated by its recipient. Sequentially unpicking the ways in which participants orient to each other's social interactions is fundamental to the ways in which claims are made and evidenced in CA.

2.3.2 Participant Orientations as a Measure of Significance

In addition to participant orientations being used to excavate the interactive categories of analysis and warrant claims made about the design and order of social interaction in CA research, participant orientations are also used to establish the *significance* of an observation. This is because the significance of an observation in CA is established in accordance with whether or not the participants engaged in the interaction orient to that observation as being significant: if they do not, then that feature is regarded as being *unimportant*, as the participants did not treat it as being important; if, on the other hand, the participants *do* orient to it as being significant, then that feature is considered to be *significant* regardless of how many times it may occur in the interaction. As pointed out by Haakana (2002: 228): ‘For the interactants the question is not how ... [the item under investigation] is statistically distributed, but what happens here and now in the interaction’. In other words, what is crucial in CA is not how many times device X, for example, occurs in talk-in-interaction but that X *did* occur and was treated as being significant in that instance by those speakers, as evidenced by their treatment of and orientations to it (Schegloff 1987a, 1993; Wootton 1989).

One consequence of using participant orientations as a measure of the significance of an observation is that the analysis of single cases becomes a relevant activity (see Schegloff 1987a, 1993). As argued by Schegloff (1993: 101; emphasis not added) ‘*one* is also a number, the single case is also a quantity’. This means that unlike in macro-investigations of talk, such as the Variationist studies outlined in 1.3, which are typically based on huge data collections of a given feature, CA analyses can be based on the unpicking of just *one* instance of a particular feature or phenomenon if the participants orient to that feature on that occasion as being relevant. Importantly, detailed single case analyses are not regarded as being inferior to those which boast large data collections, as the detailed explication of one instance of interaction using the methodology of CA is considered to be able to ‘yield a specification of interactional

considerations bearing on it which can *prove valuable* in shaping our ideas about the nature of particular phenomena contained within it' (Wootton 1989: 256; emphasis added). Once a particular feature of talk-in-interaction has been identified as being significant to the interactants, the analysis can then (if deemed necessary) proceed on a case-by-case investigation of single instances. As argued by Schegloff (1993: 102; emphasis not added):

'...in examining large amounts of data, we are studying *multiples or aggregates of single instances*. Quantitative analysis is, in this sense, not an *alternative* to single case analysis, but rather built on its back. We can be led seriously astray if we allow the possibility of quantitative analysis to free us from the need to demonstrate the operation of what we take to be going on in singular fragments of talk (except for those phenomena argued –no, *shown*—to be orderly *only* at an aggregate, statistical level).'

In other words, then, Schegloff (ibid) is claiming that when investigating the micro-organisation of talk-in-interaction, one must build the analysis in a qualitative manner by a series of single case analyses. Then, only once the workings of each individual case are understood may it be useful to apply quantitative analysis tools. The issue, however, of whether one should apply such quantitative tools and the possible benefits gained from doing so remains a heated topic of discussion and one which will not be further discussed in this thesis (for a recent overview of some of the issues and possible benefits of quantifying conversational data, see Haakana (2002)).

The participant-oriented, qualitative approach employed in this thesis may be criticised for being quasi-quantitative in that it employs 'informal quantification' terms such as *massively, recurrently, routinely, frequently, ordinarily, commonly* and *occasionally* (Schegloff 1993:99). Such terms are often regarded as being weak counterparts to more formal quantitative descriptions. However, rather than the differences in qualitative and quantitative research being regarded as 'simply weaker and stronger versions of the

same undertaking' (Schegloff 1993: 118), it is instead more fruitful to regard them as providing *different types* of accounts. As explained by Schegloff (1993: 118-119; emphasis not added):

'Terminology such as *occasionally* or *massively* reports an *experience* or *grasp* of frequency, not a count; an account of an investigator's sense of frequency over the range of a research experience, not in a specifically bounded body of data; a characterization of distribution fully though tacitly informed by the analytic import of what is being characterized.'

The absence of statistical measures of significance in favour of 'participant-oriented proof procedures' (Couper-Kuhlen & Selting 2001: 2) is markedly different from the above outlined Variationist approach (see 1.3 above) in which quantitative techniques are typically applied in order to establish the relevance and significance of an observation. Unlike the latter approach, in which statistical measures of significance are applied and cases which do not fit the overall pattern are typically disregarded as being unimportant and irrelevant (unless, of course, they result in the overall pattern being statistically insignificant), in CA analyses in which participant orientations are used as a measure of relevance, the cases which do not fit the pattern are *not* excluded from the analyses. Instead, they are incorporated into and accounted for within the analysis itself. This routinely results in 'deviant cases' being used to back up the analysis of the normative practice observed, as when speakers produce talk which analysts identify as being deviant, speakers routinely display their orientation to that talk as being atypical and thus a deviation from the norm. Thus rather than necessarily evidencing some kind of flaw in the analysis, deviant cases can serve to strengthen claims made about a normative or canonical practice (see e.g. Schegloff 1968, 1998; Selting 1998; Wootton 1989). However, deviant cases can also indicate a shortcoming in the original analysis of the normative practice and can therefore highlight a need for a new analysis in which the deviant cases can be included (see Schegloff 1968 in which he had to redo his analysis of 499 tokens in order to incorporate one deviant case).

In summary, this section has focused on the data-driven approach of CA, since its empirically grounded methodology underpins the (interactional) research in this thesis. Some of the key assumptions held in CA have been outlined and the typical domains of evidence appealed to in CA investigations of talk-in-interaction have been presented. Data fragments have been used to exemplify the import of participant orientations when constructing and evidencing arguments about the sequential design of social interaction in CA and the use of participant orientations as a measure of significance was also presented. The next section (2.4) will examine the phonetic approach adopted in the analysis chapters in 3 to 5.

2.4 Phonetic Approach

Alongside the sequential techniques of CA outlined above, this research employs both impressionistic and instrumental phonetic analysis techniques. Impressionistic techniques were given primacy in order to reflect the fact the acoustic channel was the only information present to the participants at the time at which the telephone interactions occurred. However, instrumental techniques were employed to conduct repeated listenings to the corpora and to obtain evidence, where possible, to support the impressionistic observations. Such evidence took the form of, for example, pitch and amplitude measurements. These two techniques—impressionistic and instrumental—will be outlined below in 2.4.1 and 2.4.2 respectively.

2.4.1 Impressionistic Analysis

The technique of ‘analytic impressionistic observation’ was developed by John Kelly and John Local at the University of York (see e.g. Kelly & Local 1989a, 1989b) and has since been adopted by numerous scholars wishing to explore the phonetic (and interactional) details of social interaction (see Ford & Couper-Kuhlen 2004: 11-14, for a summary of this approach to phonetic analysis). Underpinning this method of phonetic

observation is the view that too much phonetic detail in phonetic records of language is impossible, as one cannot know from the onset of an exploration which factors will turn out to be significant (Kelly & Local 1989a: 26). This impressionistic approach therefore involves observing and recording as much detail as possible about the language data under investigation with no preconceived assumptions about what will be relevant (as these may influence the documenting of particular phonetic events).

In order to record as much phonetic detail as possible, the technique of impressionistic observation emphasises the need for two types of phonetic observational techniques: kinaesthetic awareness and analytic parametric listening (see Kelly & Local 1989a, 1989b). *Kinaesthetic awareness* refers to the technique by which the analyst consciously attempts ‘to construe the heard material as audible vocal gestures’ and therefore replicates not only the sounds but also ‘the movements which give rise to those sounds’ (Kelly & Local 1989a: 29). After reproducing a person’s organic movements and sounds the investigator can then, through detailed introspective analysis, seek to establish the component parts of the production and how and where they occur in the vocal tract (ibid: 30). *Analytic parametric listening* concerns the technique by which the analyst listens to and records various phonetic parameters separately from each other (Abercrombie 1965; Kelly & Local 1989a, 1989b). This means that parameters such as, for example, pitch and voice quality are listened to separately and their details recorded independently rather than being conflated. The technique of analytic parametric listening can therefore help the investigator ‘to untease the various strands of the auditory sensation into component factors’ (Kelly & Local 1989a: 30) and to steer away from some of the prejudices that can arise from knowing one’s own language, the latter of which was a factor in this research.

Following Kelly & Local (1989a, 1989b), the data in this research were approached with as few preconceived assumptions as possible about what would be found to be relevant and significant. This approach therefore aligns with the CA maxim that ‘no

order of detail in conversational interaction can dismissed a priori as disorderly, accidental or interactionally irrelevant' (Heritage 1989: 22). Approaching the data in this way enabled previously unobserved patterns in the phonetics-interaction interface to be identified, such as the observation that clicks can engage in interactional work in talk-in-interaction.³ The parametric listenings revealed a number of phonetic parameters of importance for the interactants which included pitch, loudness, voice quality and articulatory features. Moreover, these parameters were found to cluster in different ways according to their sequential and interactional environments, as discussed in chapters 3 to 5. The technique of kinaesthetic awareness enabled uncertainties in various articulations, such as those produced in the clicks, to be clarified.

In the analysis chapters, some articulatory observations are presented using the International Phonetic Alphabet (1999), as shown between square phonetic brackets. These transcriptions are typically enhanced by the inclusion of particular sequential components of the interaction such as pauses and inbreaths. All of these additional features are denoted using courier font, as this font is used on all the interactional transcriptions. In the plain CA transcriptions of the interaction, clicks are shown using IPA conventions rather than CA notations. This means that items typically shown as .p or .t in CA conventions, which are usually deployed to capture bilabial or alveolar clicks, are instead notated as [⊙] or [!] respectively.

2.4.2 Instrumental Analysis

The speech analysis freeware PRAAT was used to listen to the speech data and relevant observations were extracted, along with their surrounding contexts, and copied into

³ Other work has also identified orders of phonetic detail in naturally-occurring speech although these works have not focused on the relationship between the patterns observed and the interactional design of the talk-in-interaction (see e.g. Simpson 1991; Shriberg et al. 1998; Foulkes & Docherty 1999; Hawkins 2003; Shriberg et al. 1998).

separate AIFF files (Boersmal & Weenink 2005).⁴ Detailed impressionistic and instrumental analyses of these fragments were then undertaken, which revealed a number of phonetic parameters of importance for the interactants. These parameters included pitch, voice quality and loudness. This section will document how the instrumental measurements of these parameters were undertaken.

a) Measurement of Pitch

Pitch was measured using the default “periodicity-to-pitch” function in PRAAT.⁵ This function attempts to calculate the fundamental frequency (F_0) for each time slice of the speech signal, in order to produce a pitch trace. However, the algorithm cannot infallibly generate a definitive pitch measurement for each time slice and a set of candidate pitches, each with an associated likelihood, is therefore computed. The program then selects the most likely candidate pitch measurement of the fundamental frequency in Hertz (Hz). Where a wrong pitch is identified as the fundamental, which is typically caused by creaky voice or an inaccurate detection of voicing in the speech signal, a different candidate pitch (or no pitch) is selected by hand after auditioning the alternative pitch traces in tandem with the speech fragment being investigated. This process is undertaken for each time slice on each pitch plot until the pitch trace derived from measurements (in PRAAT) and the pitch of the speech fragment are perceived as being ‘identical’.

b) Choice of Pitch Representation

⁴ PRAAT can be downloaded from <http://www.praat.org>.

⁵ The default pitch floor/ceiling (75-600 Hz) and time slice settings (automatic) were used for this function, except in the case of one speaker (Lesley/Holt Corpus) whose pitch ceiling exceeded 600Hz. For this speaker, a modified range of 125-650 Hz was therefore used.

The above outlined (see (a)) process of pitch measurement produced absolute, linear measurements in Hz. Two changes to the representation of this data were implemented to assist the analysis. Firstly, the measurements were changed to a logarithmic (semitone) scale, as this is regarded as corresponding more closely to the human perception of pitch than a linear (Hz) scale (Couper-Kuhlen 1996; Hayward 2000). Secondly, a measurement relative to the range of each individual speaker was preferred because, as argued by Couper-Kuhlen (1996: 371), ‘absolute pitch values...are meaningless unless they are put in relation to a speaker’s individual voice range’.

In order to use a relative scale, it was first necessary to establish the absolute pitch range of each speaker. These ranges were determined by measuring the highest and lowest pitches achieved by each speaker during roughly five minutes of interaction. The absolute highest and lowest pitch placements were then used to determine the speaker’s individual range, and the semitone span of this range was calculated using a PRAAT script designed to convert Hertz to Semitones.⁶

c) Production of Pitch Plots

The pitch traces presented in this thesis were produced using a PRAAT script which plots absolute pitch data on to a speaker-specific, relative semitone scale using the following formula:

$$s = 12 \times \log (f / b) / \log (2)$$

where *s* is the semitone value, *f* is the frequency of the time slice under investigation (in Hz), and *b* is the speaker-specific baseline frequency (also in Hz). In addition to the baseline frequency, this script requires a semitone range for the speaker to be inputted in order to determine the correct scale of the pitch axis. Horizontal lines have been added

⁶ Thanks to Gareth Walker for designing the PRAAT scripts used in this research and for making them available via his personal webpage: www-users.york.ac.uk/gw115.

to the pitch plots, which approximately divide each speaker's range into three parts (low, mid and high) in order to assist with interpretation.

Figures which simultaneously show pitch data for several fragments (spoken by different speakers) have also been produced. In order to maintain pitch measurements relative to speakers' individual ranges it was necessary to abandon the semitone scale, as not all speakers have the same semitone range. Accordingly, a logarithmic percentage scale was adopted, where for each speaker, 0% denotes the bottom of their own range and 100% denotes the top. Percentage values were calculated according to the following formula:

$$\text{percentage} = \log (f / b) / \log (t / b) \times 100$$

where *f* is the frequency under investigation, *b* is the baseline frequency for the speaker and *t* is the top of the speaker's range (all in Hertz)⁷.

d) Production of Amplitude Traces

It is well known that 'amplitude is related to our sensation of loudness' (Hayward 2000: 32). However, it has been suggested that 'perceived loudness is... tied to acoustic intensity' (Johnson 1997: 36); therefore a measure of intensity was used to approximate the loudness of the speech productions. The amplitude traces were derived using PRAAT's default "to-intensity" function, which measures the intensity in decibels (dB) for each time slice. Although instrumental measures of intensity are notoriously problematic because of factors such as differences in inherent amplitudes of speech 'segments' (Ladefoged 2001) and the lack of control of the positioning of the participants to the recording device (particularly in the corpora used in this thesis), instrumental measurements were used where they adequately reflected the author's impressionistic observations.

⁷ Thanks to Crispin Cooper for his generous help with these conversions.

e) Production of Spectrograms and Waveforms

The spectrograms and waveforms were produced using the default functions provided for this purpose and the former are plotted on a linear (Hz) scale between 0-5000 Hz. Horizontal lines were added to the spectrograms at intervals of 1000 Hz. All observations about spectrograms are made only between the frequencies of 300 Hz – 3300 Hz, as the upper frequency of the audio bandwidth carried by the telephone is ‘accepted to be 3.3 KHz at best...and at the low end of the spectrum, the telephone network carries no lower than 220 Hz and most commonly only as far down as 280 Hz or 300 Hz’ (Rodman 2003; see also Künzel 2001 and Nolan 2002 for discussions on the influence of telephone transmission on the measurement of formant frequencies in speech).

In sum, this section (2.4) has documented the phonetic analysis techniques used in this thesis. These techniques comprise a combination of (1) impressionistic investigations and (2) instrumental investigations. In all cases, instrumental investigations are used only to back up the impressionistic observations.

2.5 Summary

This chapter has detailed the data and methodologies employed in the analysis chapters in 3 to 5. First the details of the corpora and the transcription conventions employed were documented. Next, the interactional methodology of CA was presented and some of the fundamental assumptions and analytic principles which comprise this sequential-interactional analysis technique were highlighted. Finally, the phonetic analysis techniques of impressionistic observations and instrumental investigations were

examined. Previous research which has combined the analysis techniques of CA and impressionistic observations, as is done in this research, has successfully demonstrated a number of systematic and recurrent interrelationships between these two levels of organisation (see e.g. Local & Kelly 1986; Local et al. 1986; Local 1992, 1996, 2004; Couper-Kuhlen 1993, 1996, 2004a, 2004b; Tarplee 1993; Auer 1996; Wells & Peppé 1996; Ford & Thompson 1996; Selting 1996, 2000; Wells & Macfarlane 1998; Fox 2001; Ogden 2001, 2003, 2004, forthcoming; Wright 2001, 2002a, 2004; Curl 2002, 2004, 2005; Jaspersen 2002; Local & Walker 2005; Curl et al. 2004; Szczepek Reed 2004; Walker 2003, 2004a, 2004b; Ogden et al. 2004). This thesis contributes to this growing body of research.

CHAPTER THREE

MULTI-UNIT, FIRST-CLOSING TURNS

3.1 Introduction

This chapter focuses on the management of telephone-call closings in talk-in-interaction. Specifically, it investigates the sequential, interactional and phonetic properties of one device which speakers regularly use in order to move the direction of their conversation from some on-topic talk, into the closing section of the call. This functional device will be referred to as a multi-unit, multi-action first-closing turn, henceforth abbreviated to *F1 turn*. It is called a first-closing turn because it initiates the closing section of a call and therefore comprises the first turn of the closing section. It is also called a multi-unit, multi-action turn, as it is minimally composed of two units of talk, each of which undertake different actions: one attends to the preceding talk and functions to close it down and a following unit initiates and offers the action of call closure.

Fragment 1 provides an illustration of a multi-unit, multi-action F1 turn, produced by Lesley in Line 4. Observe that in the first part of this turn (after the initial inbreath (.hhh)), Lesley produces *yep* (L4). This comprises the first unit and it receipts and closes down the preceding sequence of talk. Next, Lesley states *okay then* (L4); this chunk constitutes the second unit of the F1 turn and it functions to offer and initiate the action of call closure, as evidenced by Lucy's subsequent acceptance (L5-6), and the following termination of the call (L7-8).

Fragment 1: Holt.Oct88.1.8/needle/¹

01: Luc: =[could have a few people faintin[g [at the si-:ght of=
02: Les: [ehh[heh huh
03: Luc: =i-:t[eh ah [ah
04: → Les: [.hhh [yep (0.2) [⊙] (.) o:[k a[y then]
05: Luc: [.hh[Okay] Les see you
06: tomorrow[()
07: Les: [bye:
08: Luc: bye::
- - end of call - -

The examination of F1 turns in this chapter is important for two reasons. Firstly, it exemplifies an analysis which begins from interactional function rather than linguistic form, the latter being the typical starting point in most linguistic investigations. Instead of being comprised of linguistically identical items, the collection of F1 turns is therefore comprised of interactionally identical items, as all the F1 turns occur at the same location in sequence and undertake the same interactional functions: closing down one sequence, and initiating and offering a new sequence; that of call closure. The second reason why this chapter is important is that clicks are regularly and systematically produced in the boundary between the two units which comprise the F1 turns. This can be seen in fragment 1 above in which there is an inter-unit bilabial click (transcribed using the International Phonetic Alphabet as [⊙]). The routine occurrence of clicks in this identical sequential location therefore suggests they are a component of talk which is under speaker control.

The organisation of this chapter is as follows. First, in this introductory section (3.1), an overview of previous research conducted on call closings is provided (3.1.1) and the study is outlined (3.1.2). Next, a detailed analysis of the interactional characteristics of the F1 turns in the collection is presented (3.2). Then, some of the typical lexical properties of F1 turns are presented (3.3) followed by a discussion of their phonetic characteristics (3.4). Finally, the study is summarised (3.5).

¹ In this chapter the pseudonym name of ‘Joyce’ has been changed to ‘Lucy’ in order to better reflect the disyllabic structure of the original name.

3.1.1 Previous Research on Call Closings

Research into the interactional organisation of telephone call closings (see for e.g. Schegloff & Sacks 1973; Davidson 1978; Button 1987, 1990a, 1991) is relatively sparse when compared with the vastness of research conducted on its counterpart, that of call openings (see for e.g. Schegloff 1968, 1979b; Godard 1977; Sifianou 1989, 2002; Hopper & Doany 1989; Hopper et al. 1991; Houtkoop-Steenstra 1991; Hopper 1989, 1992; Luke 2002; Park 2002; Taleghani-Nikazm 2002). Part of the reason for the general lack of research into call closings can be attributed to their complexity and elusiveness as, unlike call openings, call closings can span many turns at talk and after being initiated they need not be completed resulting in a potentially infinite number of movements in and out of closing implicative environments in one ‘unit’ of conversation (Schegloff & Sacks 1973: 292).

Not surprisingly, then, much of the previous research on call closings concerns the ways in which conversationalists move *out* of closing implicative environments (Schegloff & Sacks 1973; Button 1987, 1990a, 1991). Other research has investigated the interface between the organisation of closings and other factors such as, for example, child language acquisition (Greif & Gleason 1980), second language learning (Bardovi-Harlig et al. 1991), cross-cultural differences (House 1982; Aston 1995; Placencia 1997; Pavlidou 1997, 1998, 2002), communication impairment (Lesser & Milroy 1983; Collins et al. 1997) and psychology (Clark & French 1981). Far less attention has been given to identifying and unpicking the various ways in which parties-to-talk move *into* the closing section(s) of naturally-occurring telephone interaction; practically no research has examined the phonetic characteristics of such movements or indeed of call closings in general. Consequently, to date, all that is known about the phonetic properties of movements into closing implicative environments is what has been stated by conversation analysts; namely, that first-closing turns comprised of items such as *okay* can be produced with a falling pitch contour (Schegloff & Sacks 1973: 303;

Goldberg 2004: 259f) or with a rising pitch contour in order to check whether the preceding topic is complete and thus that a movement into closure is licensed (Button 1990a) (but see the discussions by Auer 1990, Button 1990b and Auer 1992 on the relevance of rhythm in German telephone closings; Douglas-Cowie & Cowie 1998 and Couper-Kuhlen 2004b on the various exploitations of pitch at different sequential locations in talk, including closings; and Goldberg 2004 on amplitude shifts in closings). The research in this thesis seeks, in part, to rectify these gaps in knowledge, as it provides a detailed sequential investigation of both the interactional and phonetic properties of F1 turns in talk-in-interaction.

The collaborative closing down of a telephone conversation is regarded as being an ‘interactional achievement’ (Schegloff & Sacks 1973: 290). This is because in order for a conversation to be *collaboratively* brought to close, the participants must reach a point where the relevance of speaker transition, which is an integral and recurrent property of the turn taking system (Sacks et al. 1974), is suspended. The word ‘collaboratively’ is italicised in the preceding sentence in order to highlight the fact that a telephone-call can ultimately be closed down at any point by one of the interactants putting the phone down. However, such an abrupt manner of call closure is not typical in everyday telephone interaction. Rather, people are routinely faced with the interactionally delicate task of how and when to negotiate moving into the closing section of a call in order to achieve successful, collaborative call closure.

In their seminal paper on call closings, Schegloff & Sacks (1973: 295) examine how it is possible for speakers to ‘coordinate the suspension of the transition relevance of possible utterance completion’ (Sacks et al. 1974). They argue that in order for participants to reach a point ‘where one speaker’s completion will not occasion another speaker’s talk, and that will not be heard as some speaker’s silence’ (Schegloff & Sacks 1973: 294-295), two crucial components are required:

‘...*the terminal exchange* which achieves the collaborative termination of the transition rule, and *the proper initiation of the closing section* which warrants the undertaking of the routine whose termination in the terminal exchange properly closes the conversation’ (Schegloff & Sacks 1973: 318; italics added).²

These two crucial components of call closure put forward by Schegloff & Sacks (ibid) can be seen in example 1. The first ‘okay’ in Line 1 (L1) constitutes the first crucial component, that of the ‘proper initiation of the closing section’ (Schegloff & Sacks 1973: 318). It initiates, offers and makes relevant the action of call closure. Turns which perform this action in call closure are referred to as ‘pre-closing turns’ (Schegloff & Sacks 1973: 303) and/or, more recently, ‘first-closing turns’ (Button 1987: 101). In the subsequent turn or ‘second closing turn’ (Button 1987: 100), the recipient then accepts (aligns with) the offer to close by producing a confirming *okay* (L2). Note that this second closing turn not only functions to accept the offer of call closure implemented in and through the preceding first-closing turn but it also serves to warrant the subsequent production of the second crucial component of call closings, that of the ‘terminal exchange’ (see L3-4) (Schegloff & Sacks 1973: 318). After the production of the terminal exchange, a place is marked at which a ‘collaboration on [the] termination of the transition rule can be located’ (ibid: 322), resulting in the (optional) closing down of the talk. It is after the production of the terminal exchange that the relevance of speaker transition is suspended (Schegloff & Sacks 1973; Button 1987, 1990a, 1991).

Example 1: taken from Schegloff & Sacks (1973: 317)

01: A: O.K.
02: B: O.K.
03: A: Bye Bye
04: B: Bye

² Schegloff & Sacks (1973) comment that at any stage within and a short time after the production of these two crucial components, the talk can move out of the closing implicative environment.

This four-part design of the closing section of a call, comprising a two-turn sequence of ‘offer + acceptance’ (as in *okay-okay*; see L1-2) followed by a two-turn sequence of terminal exchanges (as in *bye-bye*; see L3-4), can be divided into a sequence of two ‘adjacency pairs’ (Schegloff & Sacks 1973: 295). An adjacency pair (henceforth AP), according to Schegloff & Sacks (1973: 295-296; see also Sacks 1992b; Schegloff 1995), is the basic unit which comprises constructions of sequences in talk-in-interaction. In their most basic form, APs are composed of two turns-at-talk, which occur one after the other and are produced by two different speakers. Not only are the two turns typically placed adjacently to each other in an AP, but they also have a special status in that they constitute a ‘pair of actions’ (Heritage 1984a: 245). This is because the action undertaken in the first turn is recognised as being the first pair part (FPP) to the action undertaken in the second turn, which is recognised as being the second pair part (SPP) to the first. It is on the completion of the SPP that the action sequence embodied in the AP is completed. In call closing sequences, then, the first AP of the closing section ordinarily contains the initiation and offer of the closing section and its acceptance (as in L1-L2 in example 1). The second and subsequent AP typically contains the terminal exchange (see L3-L4), which also conforms to the AP requirements outlined by Schegloff & Sacks (1973: 295-296). This two AP construction has generally been accepted as being the archetypical framework of call closing sequences (see e.g. Jefferson 1973; Davidson 1978; Button 1987, 1990a, 1991; Bardovi-Harlig 1991; Aston 1995; Collins et al. 1997; Placencia 1997; Pavlidou 1998, 2002; Antaki 2002).

3.1.2 Outline of Study

The purpose of the study in this chapter is to examine the sequential properties of F1 turns in talk-in-interaction, as their interactional and phonetic characteristics have received relatively little attention. It aims to identify whether F1 turns are produced with any systematic and recurrent interactional and phonetic properties, and how these properties are interlinked.

In any one telephone call, there can be numerous movements into and out of the closing section of the call. However, in this chapter only those turns in which the conversation is shifted from some on-topic talk into a ‘closing implicative environment’ (Schegloff & Sacks 1973; Button 1987, 1990a, 1991) and which actually result in the subsequent closing down of the call, are considered to be first-closing turns. Therefore, closing initiations which are followed by movements out of closure by, for example, the introduction of previously ‘unmentioned-mentionables’ (Schegloff & Sacks 1973: 304) or previously discussed topics, are disregarded. Thus in this chapter the criterion used to establish the F1 turn collection was: (1) there being no topical material produced subsequent to their production and (2) the immediately subsequent occurrence of call closure (after the closing section).

Close examination of the data set revealed that speakers regularly utter F1 turns in order to move from some topical talk into the closing sequence. These F1 turns are consistently produced with the following design: a first unit which attends to the preceding sequence and a following unit which projects forwards by initiating a new sequence which offers and initiates a new action; that of call closure. In accordance with the definition of F1 turns employed in this thesis, only those turns followed solely by closing implicative material— such as future arrangements, appreciations or the re-referencing of the reason-for-call (Schegloff & Sacks 1973; Button 1987, 1990a, 1991; Auer 1990)—and the subsequent closing down of the call, comprise the F1 collection.

A total of 24 F1 turns were identified through multiple listenings and comprise the collection studied in this chapter. These turns were produced by 11 different speakers: 6 women and 5 men, all of different ages and each with different regional accents. A large number of these turns were produced by one speaker, Lesley, arguably because she spent the most time on the phone in the corpora examined: she produced 8 of the 24 F1 turns. However, the remaining 10 speakers produced 16 F1 turns thus evidencing that these turns are not an idiosyncratic feature of one person’s speech. Each of the turns and

their sequential contexts were extracted using PRAAT and placed into separate AIFF files. Detailed sequential, impressionistic and instrumental investigations were then undertaken, following the approaches outlined in chapter 2 above. The next section (3.2) will present the interactional characteristics of the F1 turns in the collection.

3.2 Interactional Properties of Multi-Unit, First-Closing (F1)

Turns

The F1 turns exhibit and engender three key interactional properties:

- they are minimally comprised of (1) one unit which receipts and closes down the preceding sequence and (2) a following unit which initiates a new sequence and serves to initiate, offer and make relevant the action of call closure;
- the recipient of the F1 turn routinely provides a turn-initial acceptance of a movement into closure such as *okay* or *right* immediately subsequent to the F1 production;
- it is overwhelmingly the ‘caller’ (rather than the ‘called’) who produces the F1 turn.

These interactional properties are discussed and exemplified with the use of data fragments in sections 3.2.1 to 3.2.3 below respectively. It is important to note that although these different properties are discussed in separate sections, they are typical across all of the F1 data fragments. The reader will therefore observe a degree of overlap between the discussions of the fragments in the different sections. This overlap is intentional so as to strengthen the various arguments proposed concerning the sequential and interactional organisation of the F1 turns and their environments.

3.2.1 The Multi-Unit, Multi-Action Status of F1 Turns

The first key and most fundamental interactional property of the 24 F1 turns identified in the data set is that they are all comprised of (minimally) two units, each of which undertakes a distinct action: one unit performs the action of receipting and closing down the preceding sequence of talk and a following unit initiates, offers and makes relevant the subsequent action of call closure. This organisation of actions within the various units of the multi-unit F1 turns can be seen in fragment 1 (repeated and extended below) which is extracted from a call between Lesley and Lucy in which Lesley has rung Lucy to request her position in the order of some upcoming presentations to the Local Women's Institute. As with each of the data fragments discussed, the sequential details of the interaction which precedes and follows the F1 turn will be examined, in addition to the F1 turn, in order to provide a full illustration of how the F1 turn is embedded within the unfurling interactional structure of the talk.

Fragment 1 begins with Lesley's statement that she hopes her presentation on acupuncture will be interesting to people (L1), which is followed by her statement that she thinks it will be (L2), as people don't know its details (L4-5). Lucy aligns with Lesley by firstly claiming that she thinks Lesley's talk will be interesting to people (oh I'm sure it will be (L3)) and after by agreeing with her that people don't know the details of how acupuncture works (no (L6)). Lesley then continues by outlining how she was going to try and get hold of a needle (L7) but was unsure whether to as she wouldn't have known who had used it before (L9-11). The final portion of this turn is produced with laughter (L10-11), which Lucy subsequently reciprocates with laughter (L12) thereby providing support for the claim that in talk-in-interaction, laughter invites laughter (Jefferson 1979; Jefferson et al. 1987).

In her next turn, Lesley states that she has not got hold of a needle (L13), which she again terminates with laughter (L14). Lucy responds with a humorously indexed

aphoristic statement, which proposes to close down the prior topic/sequence and is also produced with laughter turn-finally: perhaps it's just as well could have a few people fainting at the sight of it (L15-19). Lesley receipts and ratifies Lucy's proposed sequence/topic closure by subsequently producing *yep* (L20), a variant of *yes*, which comprises the first unit of the F1 turn. Recall that the CA methodology employed in this research maintains that any claims made about the organisation of talk-in-interaction must be grounded in the observable orientations of the parties-to-talk rather than the analyst's own speculations and intuitions. Accordingly, the claim that Lesley's turn-initial *yep* in L20 serves to receipt and ratify the preceding sequence is evidenced in Lesley's subsequent speech production, as rather than continuing with her preceding topic, she instead pauses for 0.2 (L20) seconds and then produces *okay then* (L20).³

Lesley's *okay then* constitutes the second unit of the F1 turn and it initiates, offers and makes relevant the subsequent action of call closure. Evidence for this claim is found in the subsequent orientations of the participants, as Lucy evidences that she treats Lesley's turn in L20 as being a move to close and that she aligns with it by producing *okay Les* (L21); this chunk provides Lucy's acceptance of a movement into the closing sequence. Moreover, Lucy does not provide any further topical talk; instead she subsequently references a pre-arranged engagement with Lesley for the following day: *see you tomorrow* (L21-22). Lesley then orients to Lucy's turn as having provided acceptance of her previous offer of call closure, as she subsequently produces the first part of the terminal adjacency pair sequence (*bye* L23), which is reciprocated by Lucy (*bye* L24), resulting in the collaborative closing down of the call.

³ Note that the 0.2 second pause in L20 is not a transition relevance place as throughout this 'silence', the speaker maintains a glottal closure thus signalling that they have not finished their turn-at-talk (cf. Local & Kelly 1986; Kelly & Local 1989b; Ogden 2001). The employment of oral closures as a turn-holding device in F1 turns is discussed in detail in 3.4.4 below.

Fragment 1 (repeated and extended): Holt.Oct.88.1.8/needle/

01: Les: I hope (.) what I will (.) have to say will be interesting .hh
 02: I think it is [pe-.hhh]
 03: Luc: [O h I'm][s u r e it will be]
 04: Les: [people don't know the] ins and outs of
 05: acupuncture do they=
 06: Luc: =no[:
 07: Les: [.hhhh and I[was going to try and get hold of a needle=
 08: Luc: [w e l l I d o n 't]
 09: Les: =but then I thought .hhh we:ll I don't know perhaps I won't
 10: .hh the- (0.2) e-hhh who do I know it's been stu(h)uck in uh heh
 11: he:h hih
 12: Luc: huh huh ho ho[ho ho [ho
 13: Les: [.hhh [so I haven't got hold of a needle
 14: Les: uh huhh huh huh.[.hhhh
 15: Luc: [perhaps it's just as well=
 16: Les: =[eh
 17: Luc: =[could have a few people faintin[g [at the si-:ght of=
 18: Les: [ehh[heh huh
 19: Luc: =i-:t[eh ah [ah
 20: → Les: [.hhh [yep (0.2) [O] (.) o:[k a[y then]
 21: Luc: [.hh[Okay] Les see you
 22: tomorrow[()
 23: Les: [bye:
 24: Luc: bye::

- - end of call - -

Fragment 2 provides another instance of an F1 turn in which the two distinct actions of receipting and closing down the preceding sequence, and of initiating and offering the following (call closing) sequence, are undertaken. This fragment is from a call between Lesley and her friend Bodwin in which Lesley has called Bodwin to inform her that she will be attending the pre-arranged meeting at her place that night. The fragment begins with Bodwin's statement that she is unsure of how many people will be attending the meeting: it might be more I don't know (L1). Lesley receipts Bodwin's statement with yes (L2) and then informs her that she had intended to bring along Mrs Lamp (L2-3) from North Cadbury (L5). After receiving no acknowledgment from Bodwin after either of these turns (observe the gaps in Lines 4 and 6), Lesley then continues by stating that this lady (Mrs Lamp) cannot attend the meeting that evening due to unforeseen circumstances but that she will be come to the first meeting after Christmas (L7-9). It is at this point that Bodwin provides her first and only receipt of Lesley's informing by producing the oh-prefaced assessment oh splendid (L10). Such oh-prefaced assessments have been found by Heritage (1984b) to commonly occur at the ends of topics or news-informings.

In response to Bodwin's assessment, Lesley produces a turn initial *yep* (L11), which comprises the first unit of the F1 turn. This *yep* functions as both a confirmation of the news that Lesley has just presented and as a sequence-closing third, which proposes to close down the sequence (Davidson 1984; Merritt 1980; Heritage & Greatbatch 1991; Beach 1993, 1995). Evidence to support this claim is found in Lesley's subsequent talk, as after her *yep*, she does not initiate a new topic or continue with a preceding topic; instead she produces the second unit of the F1 turn, which is comprised of *okay then* (L11). This unit serves to initiate, offer and make relevant the action of call closure. Bodwin overlaps Lesley's second unit with a turn-initial *right* (L12) and then subsequently re-references the future meeting with Lesley: *well I shall see you* (L12). This evidences that Bodwin treats Lesley's preceding turn as initiating and offering call closure and that she accepts this offer. Then, after a slightly extended closing (L12-13), the terminal adjacency pair exchange is produced (L15-16) resulting in the collaborative termination of the call.

Fragment 2: Holt.C.1985.6/splendid/

```

01:   Bod:   It might be mo:re I don't [know]
02:   Les:           [yes:]I was going
03:           to bring Missiz La:mp
04:           (0.2)
05:   Les:   from North Cadbury
06:           (.)
07:   Les:   but she can't come because her husband's unexpectedly .hhh
08:           had to go away so she's comin::g to the f:irst one after
09:           Christmas
10:   Bod:   Oh: splendid
11: → Les:   yep (0.2) [O] okay the[n
12:   Bod:           [righ[t well I shall ]see you
13:   Les:           [see you later ]
14:           (.)
15:   Les:   B[ye bye:
16:   Bod:   [Bye

```

- - end of call - -

3.2.2 Recipient Provides Acceptance Marker

A second key interactional feature of F1 turns is that subsequent to their production, the recipients of these turns typically provide their acceptance of a movement into call closure by proffering a following turn-initial 'acceptance marker', such as *okay* or *right*.

These turn-initial acceptance markers, which are produced in the ‘second closing turn’ of the closing section (Button 1987: 100), not only display that the recipients align with a move towards call closure but also that they treat the preceding F1 turn as having undertaken the action of offering call closure. It is upon the production of the acceptance marker in the second closing turn that the terminating adjacency pair sequence is sequentially licensed and can thereby be subsequently produced, resulting in successful, collaborative call closure (Schegloff & Sacks 1973). In fragments (F) 1 and 2 above, the acceptance markers are comprised of *okay* (L21, F1) and *right* (L12, F2) respectively.

Fragment 3 contains a closing sequence in which the recipient of the F1 turn provides *okay* (L11) as a turn-initial acceptance marker of a movement into closure immediately subsequent to the F1 turn (L10). This fragment is taken from a call in which Lesley has called Lucy to ask whether she will be attending their local Women’s Institute meeting that evening. After Lucy has stated that she will not be, the talk moves to other matters which include a telling by Lesley of how rude ‘Mr R’ had recently been to her and of a lady who wants to join their Women’s Institute, as her current one is becoming too political. The fragment begins with the participants’ agreements of how they do not discuss politics at their meetings (L1-8). Then, after a micropause (L9), Lesley states *yep* (L10). This *yep* comprises the first unit of the F1 turn and it functions to receipt the preceding sequence and mark it out as being closed. Lesley then produces *okay then Lucy* (L10), which constitutes the second unit of the F1 turn and which serves to initiate and offer the action of call closure.

Lucy orients to Lesley’s F1 turn in Line 10 as having closed down the preceding sequence and offered call closure, as in her next turn she does not produce any further topical talk. Instead, she overlaps the final portion of Lesley’s naming to produce *okay then Lesley* (L11), which provides her acceptance of moving into the closing section of the call (note that the use of address terms, as in this turn, is typical in closing

sequences (Jefferson 1973). Further evidence that Lucy's second closing turn aligns with Lesley's preceding offer to close is that subsequent to a slightly extended closing (L12-15), the terminating adjacency pair sequence is produced (L15-16), resulting in the successful, collaborative closing down of the call.

Fragment 3: Holt.C.85.4/politics/

```

01: Luc: hhh Oh:: well we don't uh we're- [ we're very ]=
02: Les: [we don't di-]=
03: Luc: =[ n o : : ]
04: Les: =[discuss p]oli[tics do]we]
05: Luc: [ n o : ] no]: no
06: Les: No
07: (0.3)
08: Luc: No: (          though)
09: (.)
10: → Les: yep (0.2) [⊙] .h okay th[en Lu[cy
11: Luc: [kh]h [o:kay then[Lesley
12: Les: [see you
13: late[r
14: Luc: [we'll see you y[es
15: Les: [yes by[e:
16: Luc: [okay bye:
      - - end of call - -

```

In the data examined, there are a few deviant cases in which the recipient of the F1 turn does *not* provide an immediately following marker of acceptance of call closure. However, these instances do not result in a movement out of closure once the F1 turn, and the offer of call closure undertaken within it, has been proffered. Instead the parties-to-talk orient to the closing implicativeness of the F1 turn in their following talk by remaining within the closing section and subsequently achieving collaborative call closure. Fragment 4 provides an example of such a situation. In this fragment, the recipient of the F1 turn remains silent and does not immediately produce any talk which aligns with the offer of call closure. The F1 turn is produced by Gordon in L10: his initial *haha* comprises the first unit which serves to receipt Dana's preceding assessment (L1-7) and close down the preceding sequence of talk concerning Dana's problems with her eye; his following *okay* (L10) constitutes the first part of the second unit which functions to offer and initiate call closure (note that Gordon's offer of closure comprises *okay well uhm I shall see you uh in well whenever* (L10-15)). Dana does not immediately receipt the *initial* part of Gordon's move to close,

however, and a subsequent 0.9 second gap occurs (L11). Gordon then displays his movement into the closing section of the call by beginning to reference a future meeting (L12) which, after a further 0.5 second gap (L13), Dana eventually receipts with *yeh* (L14) (in overlap with his Gordon's production of *in* (L15)). Rather than immediately producing the first part of the terminal AP after Dana's *yeh*, Gordon orients to her delayed lack of acceptance of closure by producing the word *okay* with a final rising pitch contour (L17); this token serves to check that movement into closure is warranted. Dana responds, partially in overlap, with a confirming *right* (L18).

An additional display by Gordon that Dana has not fully licensed the closing down of the call is found in his next turn (after the 1.2 second gap (L19)) in which he produces the FPP of the terminal exchange with a rising pitch contour (*bye* L20). In this location this FPP, though designed with the lexis of bidding farewell, is not actually undertaking this function. Instead, it is checking that call closure is a relevant next action by making relevant a possible non-acceptance. This claim is evidenced in the subsequent talk. Observe, for example, that in response, Dana accepts this offer of closure by subsequently producing the SPP of the terminal exchange (*bye* L23). However, the call is not terminated at this point. Instead, Gordon produces a *further* farewell bid but this time, without a rising pitch contour (L25). That Gordon produces the FPP of the closing section again but with a non-rising pitch contour, and that this results in call closure evidences the suggestion that his preceding production of *bye* (L20) was not functioning as a FPP in the terminal exchange but rather as a pre to the terminating AP. This indicates that Gordon orients to Dana's turn in L23 as being not a SPP but rather a FPP in the terminating sequence, which licenses the subsequent production of the terminating SPP (*bye* L25), resulting in call closure.

Fragment 4: Holt.88.U.1.5/eye/

01: Dan: [No: Uhm the other eye is fine and this
02: one's (0.3) okay it just i[tches
03: Gor: [hmhhhh hmh
04: Dan: and so .hhh I suppose it's like driv[ing with
05: Gor: [hmhhhhhhhhhh
06: Dan: =hay fever=only only in one eye .hh a:nd probably not
07: [as bad because I haven't got a cold
08: Gor: [hmhhh
09: (0.4)
10: → Gor: **haha (0.5) [!] .hhhhh okay .h well uhm**
11: (0.9)
12: Gor: I shall see you (0.3) uh:
13: (0.5)
14: Dan: [yeh]
15: Gor: [in:] (0.6) Well whenever
16: (0.4)
17: Gor: ok[ay
18: Dan: [right
19: (1.2)
20: Gor: [Bye:]
21: Dan: ([])
22: (0.2)
23: Dan: bye:
24: (0.2)
25: Gor: bye:

- - end of call - -

There is one instance in the F1 collection in which some topical talk is produced by the recipient of the F1 turn *after* the F1 turn has been proffered. This deviant case is provided in fragment 5 below, which is taken from a call between Lesley and Marcia. Observe in this fragment that the F1 turn is produced by Lesley in L4: the initial unit, which is comprised of *yes*, serves to receipt Marcia's preceding declination (L3) of Lesley's offer (L1-2); the second unit, which is composed of *okay* then Marcia, offers and initiates a move to call closure. This *okay* initiated move to close is not immediately accepted by Marcia, however, as in her subsequent turn, she instead produces a humorous conclusion to the preceding topic: (he said) that'll be alright he won't be going anywhere (L7). Unfortunately, the initial portion of this turn is almost inaudible but in the author's opinion, Marcia is saying *he said*. If she does indeed initiate her turn with *he said* then this suggests that this topical stretch is prompted by some talk from a third party, namely her partner, and that it is therefore reported speech. Evidence to support this suggestion is found earlier in the call, as in response to Lesley asking if Marcia's partner would like some dinner the next day, Marcia actually asks her partner directly (and receives his response) without

increasing the loudness of her talk, thereby indicating that he is in close proximity to Marcia. If Marcia's turn is in fact reported speech then this may account for why she does not produce the acceptance marker in this location, as she is instead informing Lesley of what her partner had just uttered. Moreover, regardless of whether Marcia's talk is reported speech or not she orients to being in a closing implicative environment, as after Lesley's additional closing implicative turn (*I look forward to seeing you* L9), Marcia then proffers the turn-initial acceptance marker *okay* (L10). In other words, subsequent to her turn in L7, Marcia produces no more topical talk. Note that her subsequent production of *okay* (L10) after her 'post-F1-turn topical stretch' (L7) also supports the claim that this topical stretch is reported speech, as after she has reported the speech she then orients to being in a closing implicative environment by producing her acceptance marker of a movement into call closure.

Evidence that Lesley's F1 turn in L4 was in fact an attempt to close the call is found in her subsequent talk, as she does not take up the topical talk proffered by Marcia in L7. Instead, she minimally receipts Marcia's humorous conclusion with laughter (L9) and then re-references and shows her appreciation of a future meeting (L9), which is a typical feature of call closings (Button 1990a; Auer 1990). It is after this future referencing appreciative turn that Marcia provides her acceptance of a movement into closure with *okay see you tomorrow* (L10). Lesley orients to this turn as being the second closing turn to her preceding F1 turn and therefore warranting call closure, as she then subsequently produces the FPP of the terminal exchange: *bye bye* (L12). This move is reciprocated by Marcia (L13), resulting in call closure. The sequential organisation of this call closing indicates that even though the F1 turns may not be immediately acknowledged or accepted by the recipient, the subsequent talk, both from the initiator of the move to close and from the 'abstainer', demonstrates their orientation to its closing implicativeness.

Fragment 5: Holt.SO.88.(II).1.1/mini/

01: Les: bear in mind I have got the mini so if he'd ra::ther have it the
02: other way roun:d u- uh[m]
03: Mar: [he'll be- that'll be fi:ne=
04: → Les: yes:: (.) okay then Marcia
05: (0.2)
06: Les: [. h h h]
07: Mar: [(he said)] that'll be alright he won't be going anywhere
08: (.)
09: Les: he he he [.hhh I look]forward to seeing yo[u hh .h h]
10: Mar: [(he he he)] [okay see you] tomorrow
11: (0.2)
12: Les: bye by[e:
13: Mar: [()]
- - end of call - -

3.2.3 Caller Typically Produces F1 Turn

A third interactional property of F1 turns is that it is the person who made the call, in other words the 'caller', who routinely produces the F1 turn (this is the case for all of the above fragments, except fragment 5). This finding therefore suggests that there is a preference for callers to initiate the closing section of a call, a finding which corroborates previous claims in the literature (Sacks 1992b; Antaki 2002). Supporting evidence for this claim is that, as argued by Sacks (1992b: 366): 'if it's the caller's business to do this [i.e. to close], one has a simple explanation of why callers feel lousy when called does the closing'.

However, if there exists a preference or an 'obligation falling on the caller to initiate closings' (Antaki 2002: 21), this raises the question as to what the sequential differences are if the 'called' initiates the closing. Antaki (2002: 21) claims that the instances in his data in which the called initiates the closing section can be accounted for by the length of the phone call. He argues that in very long calls between intimates 'the caller's obligations to take the initiative in leaving are loosened to the point where closing (and marking closing) can fall, without negative implication, to either party' (ibid: 21). In the F1 collection, there are very few fragments in which the called instigates the movement into the closing section of the call. However, these do not fit the claims made by Antaki, as the initiations of the closing by the called are not situated in long calls between

intimates. Instead, the calls are less than three pages long (in transcript terms) and can be very formal in that the participants use titles and surnames to address themselves and each other. Moreover, no interactional differences could be identified between the calls in which the caller or the called instigate the move into the closing section. These findings therefore suggest that F1 turns can be used as a device to offer call closure by either the caller or the called party without any dispreferred interactional implications. Further research is needed in this area.

3.2.4 Summary

In this section the interactional properties of the F1 turns have been presented. Three recurrent and systematic interactional characteristics of F1 turns have been identified:

- they are each comprised of at least two units which undertake different actions: the first attends to and closes down the preceding sequence whereas the second initiates a new sequence in which the action of call closure is offered;
- the recipient typically acknowledges and aligns with the offer of closure by proffering a turn-initial acceptance marker (of a movement into the closing section) in their turn immediately subsequent to the F1;
- the caller routinely produces the F1 turn rather than the called.

The next section (3.3) will examine the lexical characteristics of the F1 turns.

3.3 Lexical Properties of Multi-Unit, First-closing Turns

This section will outline some of the typical lexical properties of F1 turns. In particular, it will show that:

- the first unit is typically comprised of one lexical item, usually *yes* or a variant thereof;
- in the second unit, *okay* is routinely situated as the first lexical item and is frequently followed by *then*, as in *okay then*.

These lexical characteristics will be examined in sections 3.3.1 and 3.3.2 respectively.

3.3.1 First Unit Typically Comprised of One Lexical Item

The first typical lexical property of F1 turns is that they are routinely produced with only one lexical item comprising the entirety of the initial unit. Moreover, there is an overwhelming tendency for these one-worded initial units to be composed of *yes* or a variant of *yes* such as *yep*, as is the case in all of the above fragments (except fragment 4). The F1 turns in the above fragments have been provided below for ease of reference (and the initial units emboldened):

Fragment 1: **yep** (0.2) [⊙] (.) o:k ay then (L4/20)

Fragment 2: **yep** (0.2) [⊙] okay then (L11)

Fragment 3: **yep** (0.2) [⊙] .h okay then Lucy (L10)

Fragment 5: **yes::** (.) okay then Marcia (L4)

Fragment 6 below provides another instance of an F1 turn in which the initial unit is comprised of *yep*, a variant of *yes*. This fragment is taken from a call between Mr F and Jenny in which Mr F has called Jenny to reassure her that her son, Mathew, who is camping in Mr F's garden that night with Mr F's son, will be looked after well. The transcript begins with Mr F and Jenny establishing that Mathew has his key (L1-5), which is important for Jenny, as she is concerned she may not hear him when he returns (L6). After a 0.2 second gap (L7), Mr F then produces an F1 turn (L8). Observe that the initial unit is comprised of *yep*, and it serves to receipt Jenny's prior talk and close

down the preceding on-topic sequence regarding Jenny's son. Then, after a 0.2 second pause and an inbreath, Mr F produces the second unit of the F1 turn, comprised of *okeydokey anyway* (L8); this second unit functions to initiate and offer call closure. It is accepted by Jenny with *okay* (L9) and after a show of appreciation (L9-12), the terminating adjacency pair sequence is produced (L12-13), resulting in the termination of the call.

Fragment 6: Rahman.A.1.6/okidoki/

```

01:   MrF:   do you have a ke:y uh Mathew
02:           (1.0)
03:   MrF:   yes he's go[t a key
04:   Jen:           [he's got the key with him that's alright=
05:   MrF:   =[yeah
06:   Jen:   =[just in case I don't hear him if he knocks::
07:           (0.2)
08: → MrF:   yep (0.2) [.hhh o:keydo[key a n y w a y]
09:   Jen:           [Yes          [okay well tha]nk you very much for
10:   having them anyway=
11:   MrF:   =yes that['s okay[then
12:   Jen:           [.hhh          [hhuh'kay .hhh Bye bye:
13:   MrF:   right bye now
                                     - - end of call - -

```

Previous research has claimed that in talk-in-interaction, speakers can produce the *yep* variant of *yes*—that is, the variant with syllable-final lip closure—in order to index that they will not provide any further topical talk (Heritage & Sorjonen 1994 13ff: 25; Raymond 2000: 184; see also Bolinger (1946: 90) for comments on *yep* being a ‘gesture of finality’). This is the case in F1 turns, as *yep* frequently comprises the initial unit of the F1 turn, which serves to receipt and close down the preceding talk. Other research has identified that speakers can also indicate their upcoming non-elaboration from the *onset* of their response token by producing the entire token with closed lips, as in the particle *mm* (Gardner 1997, 2001).⁴ Such particles were found to constitute the entirety of the first unit of some F1 turns, as in fragment 7 below, which is taken from a call between Lesley and Dana, her son Gordon's close friend, in which Lesley has called Dana to ask her whether she could return a book that Lesley had lent her. The talk then moves to a discussion of the people from the village who have recently moved away to

⁴ Raymond (2000: 184) claims that the response token *mm* ‘appears to be distinctly British’.

University. The fragment begins with Dana's assessment of how it is getting complicated to keep track of all the letters that she needs to write (L1). She concludes that she needs to make a list (L5-6). Lesley responds with a 'joking solution to the problem'⁵—you'll have to join them somehow join one of them and then you've only got eight to write to (L7-8)—which is received by Dana with *yeah* (.) something like that (L9-11).

In overlap with the final portion of Dana's receipt, Lesley produces a lengthened *mm*: : : token (L12), which comprises the first unit of the F1 turn. This *mm* token serves to minimally receipt Dana's preceding talk whilst also indicating that Lesley has no more topical talk to contribute (Gardner 1997, 2001). As argued by Gardner (1997: 135), a third turn *mm* receipt token is 'retrospective and sequence-closure relevant, at the same time paving the way for the introduction of a new topic'. Dana orients to Lesley's *mm* as being sequence-closing as, in overlap with Lesley's subsequent click and inbreath, Dana produces *yeah anyway* (L13), which indexes a shift in the topical direction of the talk (Drew & Holt 1998). Lesley then produces the second unit of the F1 turn, comprised of *okay then Dana*, which Dana orients to as being a move to close the call, as she subsequently provides a show of appreciation to Lesley for having phoned: *thank you very much for ringing* (L14). Thus although Dana's turn does not have a turn initial 'acceptance marker' *per se*, her show of appreciation to Lesley for having called functions in the same way, as it indicates that a movement into call closure is an acceptable next activity; such longer shows of appreciation appear to be in system with the one-worded acceptance markers. After Dana's appreciation (L14), the terminal adjacency pair sequence is produced (L15-16), resulting in collaborative call closure.

⁵ This is a comment by Gail Jefferson noted on the transcript.

Fragment 7: SO.88.(II).2.4/dana/

01: Dan: it's getting really [compli]cated [(though)
02: Les: [.hhh] [()]
03: (.)
04: Les: hhhh(h)hhh [h u h .h h h h h h]
05: Dan: [so I'm going to have to make] out a list
06: you kno[w
07: Les: [you'll have to joi:n them somehow join one of them and
08: then uh you've only got uh- eight to write to .hh[hh
09: Dan: [yeah
10: Les: [mmm
11: Dan: [something like that [hhhh
12: → Les: [mm::: (0.5) [![.hhh (0.2)]okay then Dana
13: Dan: [yeah anyway]
14: Dan: thank you very much for [ringing ()]
15: Les: [.h h h o k] ay bye:
16: Dan: () bye:
- - end of call - -

Interestingly, across all the F1 turns in the collection there is a ‘noticeable absence’ of the production of *okay* in the initial unit of the turn (see Schegloff 1995a for a discussion on ‘noticeable absences’ in talk-in-interaction). In other words, there are no occasions in which *okay* is produced in both the initial and the second unit of the F1 turn; instead, *okay* is only produced (if at all) in the second unit. This suggests that there is a sequential constraint in operation which prevents *okay* from occurring in F1 turns as both a first and a second unit component. However, this constraint may not be specific to *okay* but rather may be a generic constraint which prevents *the same speaker* from deploying the same lexical item (whatever that item may be) in sequentially adjacent units to perform different actions in talk-in-interaction. The phrase ‘the same speaker’ is italicised in the preceding sentence to indicate that this potential constraint is not in operation across *different* speakers, as there are various instances in the F1 collection in which *okay* is produced in the second unit of the F1 turn (in which call closure is offered) and is also produced by the recipient in their immediately subsequent turn (which accepts the preceding offer of call closure; see fragments 1, 3 and 6 above). Importantly, in these cases, *okay* is produced in immediately subsequent units to undertake different actions in talk and these units are produced by different speakers. These findings indicate that further research is needed in order to establish what constraints may be enforced upon the sequential placement of identical lexical items when deployed by one speaker to undertake different interactional activities in talk.

3.3.2 ‘Okay’ Routinely Situated as First Lexical Item in Second Unit and Typically Followed by ‘then’, as in ‘okay then’

The second lexical regularity in the F1 turn collection is that *okay* is routinely produced as the first lexical item in the second unit of the F1 turn. Moreover, it is frequently followed by the lexical item *then*, as in *okay then*. The high occurrence of *okay* and *then* can be seen in the F1 turns in the above fragments (1-7), which are repeated below for ease of access (the tokens of *okay* and *then* have been emboldened).⁶

- Fragment 1: yep (0.2) [☹] (.) **o:k ay then** (L4/L20)
- Fragment 2: yep (0.2) [☹] **okay then** (L11)
- Fragment 3: yep (0.2) [☹] .h **okay then** Lucy (L10)
- Fragment 4: haha (0.5) [!] .hhhhh **okay** .h well um ... (L10)
- Fragment 5: yes:: (.) **okay then** Marcia (L4)
- Fragment 6: yep (0.2) .hhh **o:keydokey** anyway (L8)
- Fragment 7: mm::: (0.5) [!] .hhh (0.2) **okay then** Dana (L12)

Fragment 8 provides an additional fragment in which *okay* is situated as the first lexical item in the second unit of the F1 turn and is immediately followed by *then*. This fragment is extracted from a call between Gay and Jeremy in which Gay has called Jeremy to update him on the current developments concerning a house that he would like to purchase. The F1 turn is produced by Gay in Line 9. The first unit, composed of the stretch *yeah I think so*, functions to receipt and agree with Jeremy’s preceding assessment, as first assessments typically project a preference for an agreeing second assessment (Pomerantz 1984). Moreover, it closes down the preceding sequence, as sequences of assessments are one device used to close down large sequences of talk (Goodwin & Goodwin 1992). Next, the second unit of the F1 turn is produced, which is prefaced with *okay* and comprised of *okay then*. This turn serves to initiate, offer and

⁶ Note that *okeydokey* (fragment 6) has been included as it is considered to be a variant of *okay*.

make relevant the subsequent action of call closure. In response, Jeremy accepts the move to close (L10) and after a slightly extended closing (L11-13), the terminal AP is produced (L13-14) resulting in call closure.

Fragment 8: Heritage 01.7/wait/

01: Gay: or unless you've got some better suggestions of getting
02: rid of Miss Ra:bins that's alright but we better not do it
03: it's the Chr[istmas spirit
04: Jer: [we:ll
05: Jer: no[:
06: Gay: [heh-heh-he[h-heh
07: Jer: [no:
08: Jer: tw- (0.2) wait until (0.7) until afterwards I think perhaps
09: → Gay: **yeah I think so <okay the:n**
10: Jer: okay(th)
11: Gay: o[kay
12: Jer: [thanks a lot a[gain
13: Gay: [not at tall bye no:w h
14: Jer: bye bye
- - end of call - -

Although *okay* occurs most frequently as the first lexical item in the second unit of the F1 turns in the collection, there are various other lexical items which are produced in this unit-initial position. These items include *alright*, *right* and *righto*. Fragment 9 provides an example of an F1 turn in which *righto* occurs as the first lexical item of the second unit of the F1 turn. This fragment is taken from a call in which Lesley has called Foster, the local vicar, to check whether there is a Sunday School the following Sunday. In the sequence preceding Foster's F1 turn, Lesley has been discussing how her family have gone skiing with Ansford school. She then proffers an assessment of her family's enjoyment of the trip—they have a great time (L4)—which, after a micropause (L5), is receipted and evaluated by Foster with *good* (L6). Foster's production of *good* comprises the first unit of the multi-unit F1 turn. It minimally receipts Lesley's preceding assessment whilst also proposing to close down the just-prior sequence. Afterwards, Foster produces the second unit of the F1 turn, which is initiated with *righto* and comprised of *righto* Lesley (L6). This second unit initiates, offers and makes call closure relevant. Lesley orients to the closing implicativeness of this turn, as she overlaps its final syllable with the acceptance marker *right* (L7). This acceptance marker indexes that Lesley has both understood and accepted Foster's preceding turn as

offering call closure. Then, after an extended closing (L8-9), the terminating adjacency pair sequence is produced (L10-11), which results in call closure.

Fragment 9: Holt.2.1/righto/

```
01: Les: they go with Ansford School
02: Fos: mm:
03:     (. )
04: Les: they have a great time
05:     (. )
06: → Fos: good (0.5) [!], hhh righto Les[ley
07:     Les: [right- ekhh he Okay [then
08:     Fos: [yes
09:     (. )
10: Les: [bye bye
11: Fos: [bye bye
```

- - end of call - -

3.3.3 Summary

This section has presented some of the typical lexical design features of F1 turns. It has highlighted that there is an overwhelming tendency for the first unit of F1 turns to be comprised of one lexical item, typically *yes* or *yep*. It has also identified that both *okay* and *then* are routinely produced in the second unit of F1 turns, as in *okay then*.

To this point, the F1 turns in the collection have been shown to be produced with a number of systematic interactional and lexical characteristics. Generally, they are comprised of two units, one which attends to the preceding sequence and one which initiates the following call closing sequence. The action of initiating call closure is evidenced in the talk subsequent to the F1 turn, as it concerns only closing implicative material and results in the successful, collaborative, closing down of the call. Moreover, not only do these F1 turns systematically undertake the same two distinct actions but they have particular lexical tendencies: the first units are typically comprised of *yep* and the second units of *okay then*. These findings therefore differ markedly from the claims in the literature that F1 turns in talk-in-interaction are typically produced by lexical

items such as *well*, *okay* and *so* which comprise the entirety of a TCU (Schegloff & Sacks 1973; Button 1990a; Goldberg 2004).⁷

In the next section (3.4), the phonetic characteristics of the F1 turns in the collection will be examined.

3.4 Phonetic Properties of Multi-Unit, First-Closing Turns

Overall, the F1 turns in the collection are produced with a number of strikingly similar and systematic phonetic properties. These properties relate primarily to the parameters of pitch, voice quality, loudness and ‘articulatory segmental’ features. The F1 turns routinely share the following phonetic characteristics:

- orderly patterns in the placement of the different units in the speakers’ pitch ranges;
- systematic differences in the relative pitch spans of the units;
- consistent patterns in the shapes of the unit-final pitch contours;
- complete closure at the end of the final syllable of the initial unit;
- an upstep in pitch in the onset of the second unit;
- glottalisation in the onset of the second unit;
- differences in amplitude between the two units;
- inter-unit clicks.

These characteristics will be examined in 3.4.1 to 3.4.8 below respectively.

⁷ The exact wording used by Schegloff & Sacks (1973: 303) is that these forms constitute the ‘entire utterance’.

3.4.1 Differences in Placement of Units Relative to Speakers' Pitch Range

One of the most marked regularities in the phonetic design of F1 turns lies in the organisation of the pitch. Both the first and second units of each F1 turn are produced with distinct pitch characteristics, which are noticeably different from each other. Furthermore, these characteristics are consistent across each of the F1 turns. In other words, the pitch designs of all the first units in the F1 turns are remarkably similar to each other, as are the pitch designs of all the second units.

A particularly salient pitch feature of F1 turns is the different placement of the first and second units in the speakers' pitch ranges: the first unit, which performs the action of attending to and closing down the preceding sequence of talk, is overwhelmingly produced much lower in the speaker's pitch range than the second unit, which initiates and offers the action of call closure. Conversely, the second unit is generally produced much higher in the speaker's pitch range than the first unit. These differences between the relative pitch placements of the first and second units can be seen in the pitch trace of the F1 turn given in figure 3.1 (which is produced by Lesley and given in fragment 2 above). Observe that the first unit, which is comprised of *yep*, is produced with a fairly low pitch placed relatively low in the speaker's own pitch range. In comparison, notice that the second unit, which is composed of *okay then*, has a much higher pitch and is produced much higher in the speaker's pitch range relative to the first unit. Such upwards shifts in pitch can be likened to 'declination resets' (Ladd 1988; 't Hart et al. 1990).

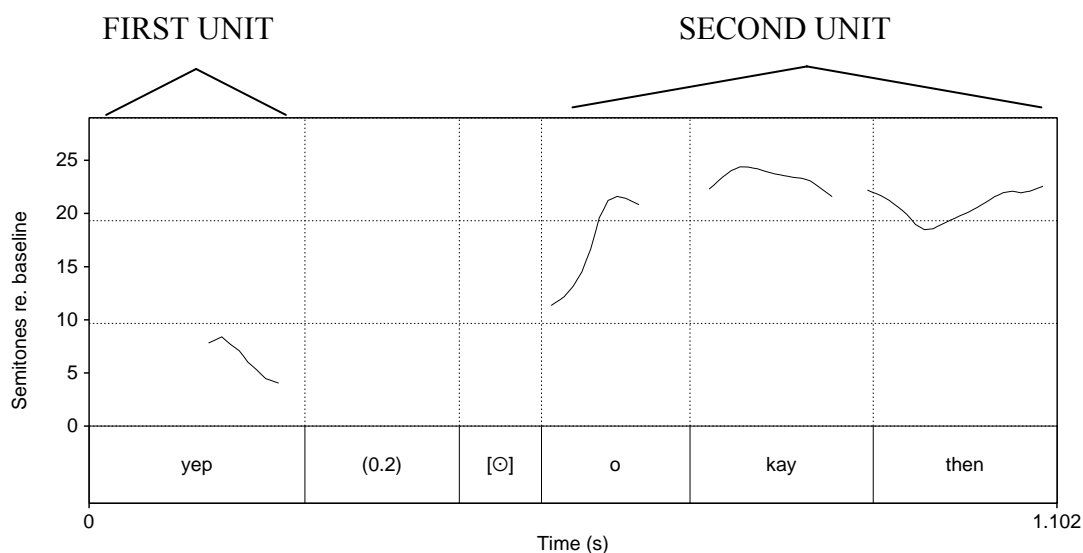


Figure 3.1: Pitch trace to show relative placements of first and second units of F1 turns (taken from Holt.C.1985/splendid/)

3.4.2 Differences in Relative Pitch Spans of Units

The second key difference between the organisation of the pitch in the first and second units of F1 turns is found in their relative pitch spans: the first unit, which attends to the prior talk, is routinely produced with a much narrower pitch span than the second unit, which offers and makes call closure a relevant next action. These differences are evidenced in the figures in table 3.1 which provides the semitone (ST) spans of the first and second units in the F1 turns in fragments 1 to 9 above. Observe that the span of the second unit is always greater relative to the corresponding first. Across the whole collection of F1 turns, then, the second unit is routinely produced with a much wider pitch relative to the first unit. Although the first units are typically much shorter in duration than the second units, the differences in relative pitch spans also hold in the F1 turns in which the first unit has a longer duration than the second unit, as in fragment 8 above (in which the first unit is comprised of *yeah I think so* and the second of *okay then*).

Fragment Number	Semitone Span of 1 st Unit	Semitone Span of 2 nd Unit
1	1 ST	14 ST
2	4 ST	12 ST
3	3 ST	16 ST
4	4 ST	7 ST
5	6 ST	19 ST
6	0 ST	8 ST
7	6 ST	17 ST
8	3 ST	9 ST
9	2 ST	5 ST

Table 3.1: Semitone spans of first and second units in F1 turns in fragments 1-9

The differences in the relative pitch spans between the first and second units can be seen in the pitch trace given in figure 3.2 below (which corresponds to an F1 turn produced by a man but which is not discussed above). Observe that the first unit, which is comprised of *yes*, has a relatively narrow pitch span of 4 semitones relative to the second unit, composed of *okay then*, which has a wide pitch span of 20 semitones.

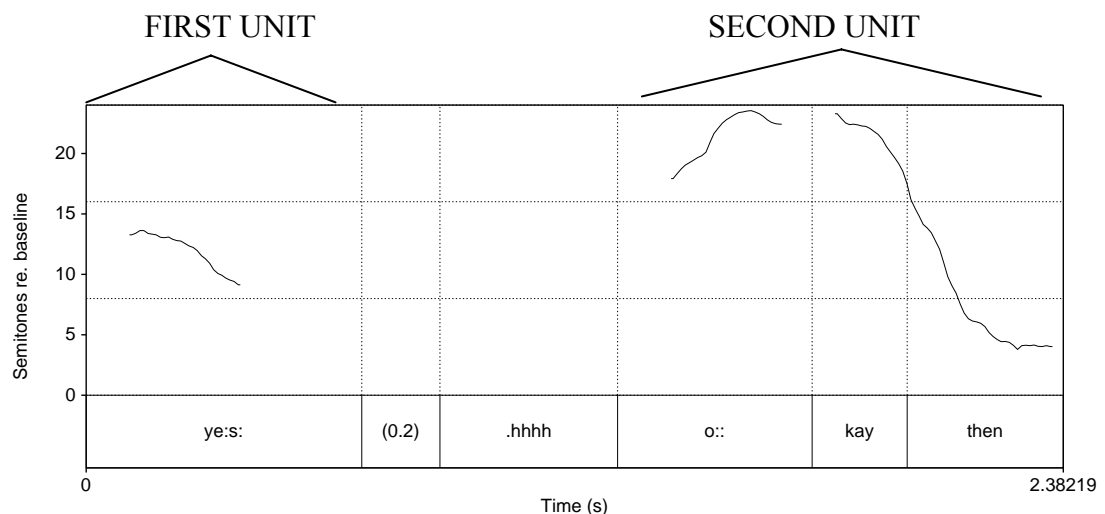


Figure 3.2: Pitch trace to show relative spans of first and second units of F1 turns (taken from Holt.88.1.11/bowling/)

These differences between the relative pitch spans and the placement of the first and second units in the speakers' pitch ranges can be clearly seen in the graph given in figure 3.3. This graph presents the pitch spans of the first and second units. It shows the highest and lowest pitches reached in each unit. Thus it can be seen that, overall, the first units are produced with a much narrower pitch span relative to the second units, as

the length of the white rectangles is generally much shorter than the black lines. This graph also shows the location of the first and second units of the F1 turns within the speakers' individual pitch ranges, as they are plotted relative to each speaker's individual range. It can be observed, therefore, that, the first units (indexed by the whitebars) are typically produced much lower in the speakers' pitch ranges relative to their corresponding second units (shown by the black lines).

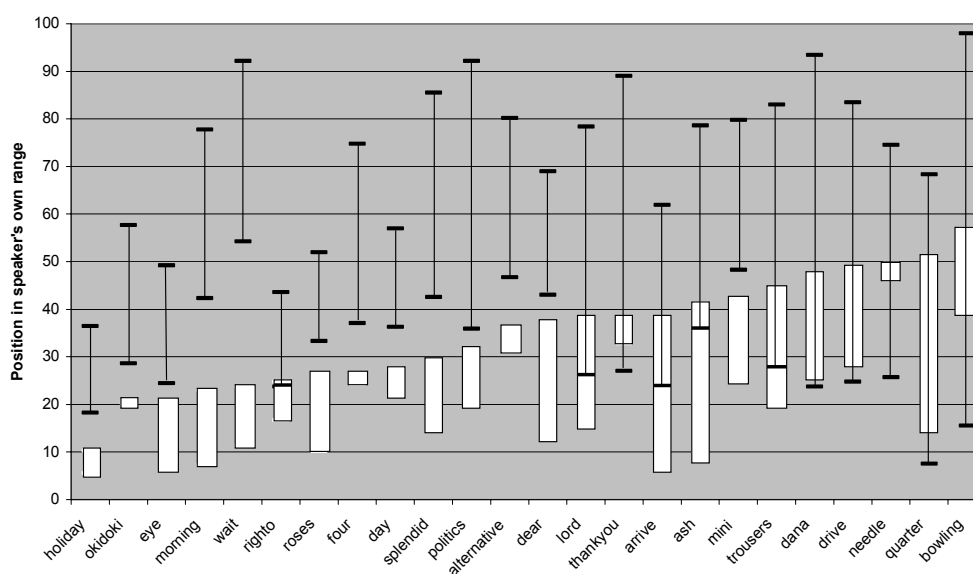


Figure 3.3: Graph to show pitch placements and spans of first and second units of all F1 turns (pitch is expressed as a percentage where 0% represents the bottom of each speaker's individual range and 100% the top; the pitch spans of the first units are shown by a white bar and the second units are shown by a vertical black line; see 2.4.2 for more details)

The consistent differences in the pitch placement and pitch spans of the first and second units of the F1 turns maps onto the two recurrent actions undertaken in these turns; namely, the first unit, which performs the action of attending to and closing down the preceding sequence is always produced lower in the speakers' pitch range and with a narrower pitch span relative to the following unit, which serves to initiate and offer the subsequent action of call closure.

3.4.3 Differences in Unit-Final Pitch Contours

The third key difference between the organisation of pitch in the first and second units of F1 turns is found in the design of their unit-final pitch contours: the first units are always produced either with a unit-final falling or flat pitch contour whereas the second units are uttered with a rising or flat unit-final pitch contour. This difference in the shape of the unit-final pitch contours is evident in the pitch trace of the F1 turn given in figure 3.4 (which is produced by Lesley but is not discussed in the above fragments). Observe in this pitch trace that the initial unit of the F1 turn, which is composed of *yes*, has a falling unit-final pitch contour whereas the second unit, which is comprised of *okay then*, has a rising unit-final pitch contour.

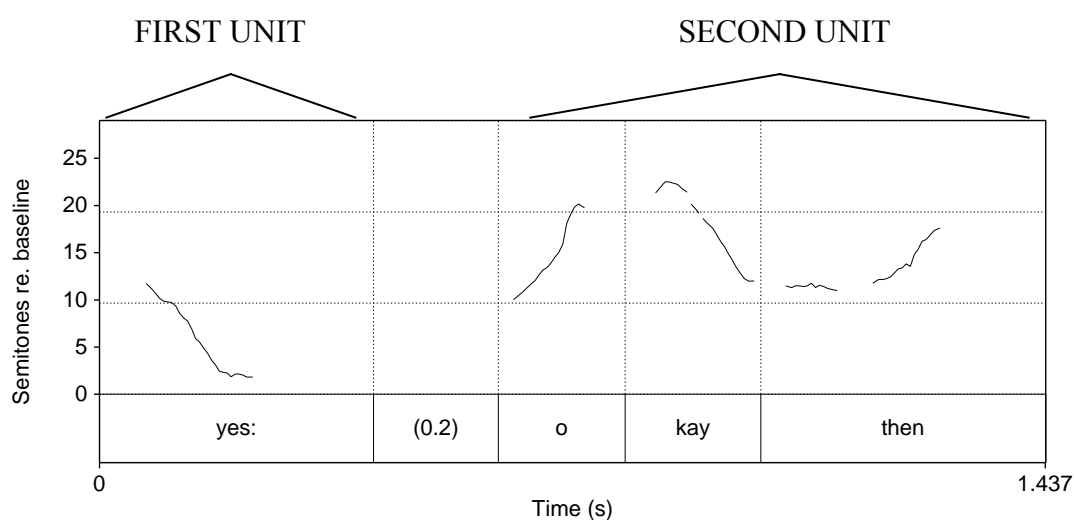


Figure 3.4: Pitch trace to show shape of unit-final pitch contour in first and second units of F1 turns (taken from Holt.2.9/holiday/)

That the final portion of the first unit of the F1 turns is consistently produced either with a flat or a falling pitch contour and the final portion of the second unit with either with a flat or a rising unit-final pitch contour suggests that there is an association between the type of pitch contour produced and the action undertaken in and through the unit. Recall that the first unit of the F1 turns performs the action of receipting the preceding sequence and marking it out as being complete. This action of closing down the prior talk thus appears to be indexed and undertaken, in part, by the lack of a unit-final rising

pitch contour. In comparison, the second unit, which initiates and offers the action of call closure, is never produced with a unit-final falling pitch contour, which again maps onto the action of the unit as, unlike the initial unit, this second unit typically seeks an acceptance/confirmation from the recipient that movement into the closing section of the call is warranted. The second unit is therefore not produced with a unit-final falling pitch contour, as it functions as a first pair part in the sequence and thereby seeks a second pair part in order for call closure to be licensed (Schegloff & Sacks 1973; Davidson 1978; Button 1987, 1990a). This action is, therefore, partly conveyed by the absence of a falling pitch contour.

3.4.4 Complete Closure at End of Final Syllable of Initial Unit

The fourth particularly salient phonetic feature of F1 turns is that a period of complete closure routinely occurs at the end of the final syllable of the initial unit. Typically, these closures occur at the end of a variant of *yes* and are overwhelmingly produced with labiality and orality, as in [jep^ʰ] although they can also occur with glottality, as in [jeʔ^ʰ]. In the cases in which the closure is produced further forward than the glottis, these closures are often ‘reinforced’ with a simultaneous glottal closure, as in [jep^ʰʔ^ʰ]. These closures are then held in the juncture between the two units and either released in the onset of a following click, an inbreath or the initiation of the second unit of the F1 turn.

In the above outlined F1 turns, four are produced with *yes* with word/unit-final bilabial closure, as in *yep* (see fragments 1-3 and 6). The phonetic properties of these F1 turns are transcribed below, although the phonetic transcriptions have been modified to include details concerning pause lengths (indexed between parentheses) and inbreaths (shown by .h). Observe in the transcriptions that the bilabial and glottal closure at the

end of *yep* is consistently held across a 0.2 second stretch of silence and then released in the onset of a bilabial click.⁸

Fragment 1: [jɪ̄p̄ ʔ̄ (0.2) ⊙ (.) ʔəʊ:kʰei]

Fragment 2: [jīp̄ ʔ̄ (0.2) ⊙ ʔkʰeʔ]

Fragment 3: [jɪ̄p̄ ʔ̄ (0.2) ⊙ .h ʔɪ̄kʰei]

Figure 3.5 provides a wave form and a spectrogram of an F1 turn in which the offset of the first unit, which is comprised of *yep*, is produced with a portion of complete bilabial closure with orality and glottality, as in [jīp̄ ʔ̄] (this is the F1 turn from fragment 2 above, produced by Lesley). Observe in the spectrogram that the onset of this *yep* token is initiated with a relatively low F1 (at around 302 Hz) and a fairly high F2 (at around 2700 Hz). This corresponds to the onset of the approximant [j], which has a low F1 and a high F2, as the tongue has a high-front position in the vocal tract (thus it is similar to the vocal tract constriction (and formant structure) for the high-front vowel [i]). After the onset of voicing in the initiation of the approximant, notice that the formants exhibit rapid transitions into the following vocalic portion; specifically, F1 rises and F2 (and F3) decrease as the speaker moves from [j] to the more open and further back vocalic gesture associated with [a]. Observe that the transition from [a] into the following bilabial closure ([p̄]) is then marked by further downwards transitions in F2 and F3, as is typical in bilabial stops. Moreover, the speaker's movement into this complete bilabial closure results in the cessation of energy at most frequencies (although a small amount of voicing from the preceding vocalic portion may perseverate into the following stretch of 'silence'). The subsequent lack of visible energy, which ensues for 0.2 seconds, corresponds to the speaker maintaining the bilabial and glottal closure.

⁸ Fragment 6 also appears to have a bilabial closure (and possibly a glottal closure) held across the 0.2 second pause (and then released in the following inbreath). However, as the recording quality of this fragment is very poor, no detailed phonetic analysis can be undertaken.

These closures are then released in the onset of the following bilabial click, which can be seen in the spectrogram (and wave form) as a short and sudden spike of energy with a frequency ranging from between 500-3300 Hz. After the click, the speaker then produces *okay then*, which comprises the second unit of the turn.

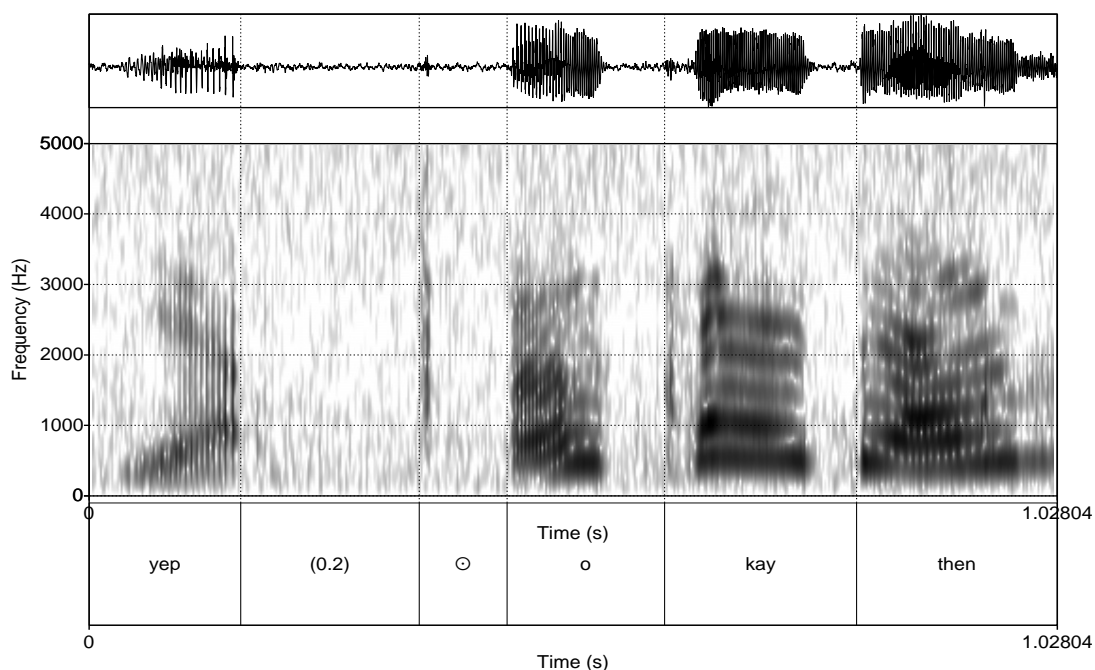


Figure 3.5: Wave form and spectrogram to show closure at end of first unit in F1 turn (taken from Holt.C.1985.6/splendid/)

The routine occurrence of *yep* in the initial unit of the F1 turns provides support for the claim that speakers employ such a manner of articulation in order to indicate that they will not further elaborate their response (Heritage & Sorjonen 1994; Raymond 2000). Such a claim is supported, as subsequent to the production of *yep*, the speakers typically do not provide any further topical talk. However, not only does this complete lip closure indicate that the speaker will not proffer any additional talk concerning the preceding topic but it also serves to index the completeness of the previous action, an observation not previously noted in the literature. In other words, the closure on *yep* in the first unit not only closes down the preceding sequence and topical talk but it also closes down the action being undertaken within and through the first unit. This closing down of an action in one unit in the multi-unit F1 turn thus enables the speaker to begin a new action in the same turn. Therefore, rather than speakers deploying *yep* 'to project

utterance completion, or no elaboration', as claimed by Raymond (2000: 194), it is employed in F1 turns to receipt and close down the preceding sequence in order that speakers *can* continue their turn and initiate an immediately subsequent new action; that of call closure.

Further evidence that bilabial closures at the end of *yes* can be deployed as an interactional resource is found in the phonetic organisation of the initial unit of the F1 turn given in the spectrogram and wave form in figure 3.6 (this is produced by Lesley and is not discussed above). Observe that the first unit of the turn, which is comprised of *yes*: ([j¹ɛs:p¹]), has a 0.25 second stretch of voicing which coincides with the approximant [j] and following vocalic portion [ɛ]. After, there is a portion of friction (as evidenced by the noisy energy distributed from 300-3300 Hertz on the spectrogram), which coincides with the production of [s] in the coda of *yes*. Then, a sudden lack of energy can be seen, which is coincident with the lips being closed and the offset of voicing. This bilabial and glottal closure is maintained for 0.2 seconds, resulting in a stretch of 'silence'. This silence is evidenced by a lack of visible energy. Then, the bilabial closure is released in the onset of a bilabial click, as indicated by the sudden high energy spike following the stretch of no energy. Immediately subsequent to this burst of energy and continuing for a period of 60 ms is a stretch of energy with high amplitude. This corresponds to the inhalation of a long and loud inbreath which is initiated as the click is released. Immediately after this inbreath the speaker then produces the second unit of the F1 turn, which is composed of *okay then*.

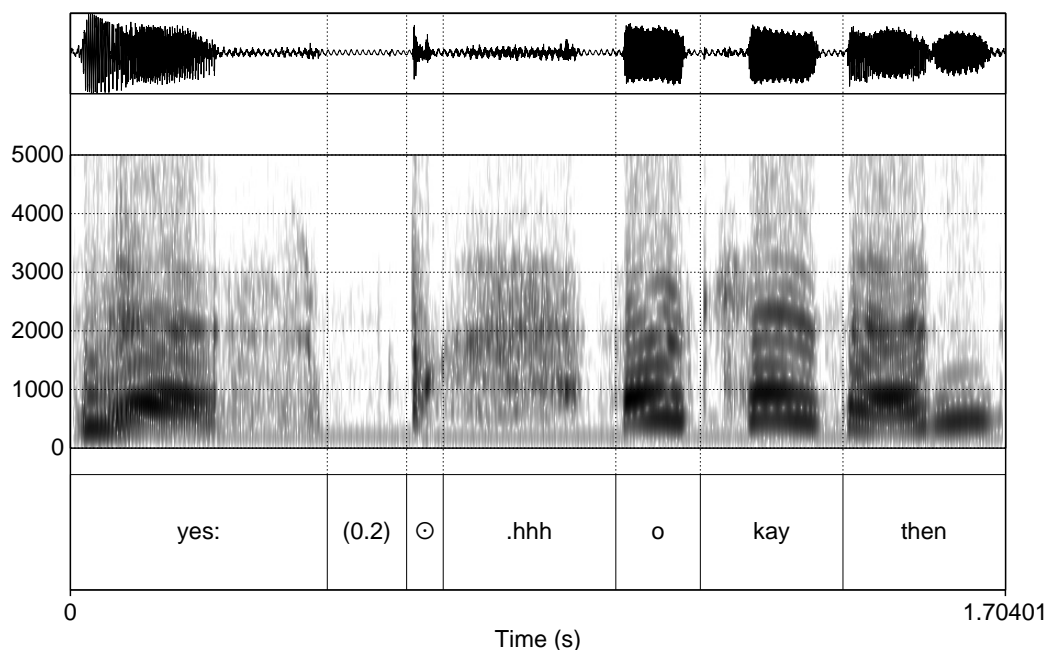


Figure 3.6: Wave form and spectrogram to show friction and bilabial closure in coda of first unit of F1 turn (taken from Holt.SO.88.1.10/ash/)

That Lesley makes a bilabial closure in the above discussed F1 turn after having already produced a stretch of friction in the final portion of *yes*, as in [jⁱɛs:p^ɾ], provides evidence against the possible argument that speakers may produce *yep* instead of *yes* as *yep* takes less articulatory effort. Instead, Lesley’s production of [jⁱɛs:p^ɾ] supports the argument that the occurrence of a period of closure in the final portion of *yes* in the first unit of F1 turns can have important sequential and interactional functions.

It is important to note that although the unit-final portions of complete closure in the initial units of F1 turns are frequently produced at the end of *yes* or a variant thereof, this is by no means the only location. Rather, they can also be produced in unit-final position (in the first units of the F1 turns) at the ends of ‘non-yes’ words, such as *no* ([nəʔ]), *mm* ([ʔm::]) and *good* ([gʊd^ɾ]). In the above data fragments, complete closure is

of the offset of the preceding first unit. In the above nine fragments, the onset of the second unit is always produced with an upstep in pitch of at least 5 semitones, as shown in table 3.2, which provides the number of semitones in the upstep in the onset of the second unit relative to the offset of the preceding unit.

Fragment Number	Number of Semitones in Upstep
1	5 ST
2	8 ST
3	15 ST
4	9 ST
5	9 ST
6	7 ST
7	7 ST
8	8 ST
9	9 ST

Table 3.2: Number of semitones in upstep in pitch in onset of the second unit in fragments 1-9

This resetting in pitch in the onset of the second unit can be seen in the pitch trace of the F1 turn given in figure 3.7 (which occurs in fragment 3 and is produced by Lesley). Observe that the first unit, which is comprised of *yep*, has a narrow pitch span and an offset which is located low in Lesley's pitch range (the offset is produced at 173 Hz hence, as her baseline is 125 Hz, it is produced only six semitones above her baseline). In comparison, the onset of the second unit, which is initiated with *okay*, is produced much higher in her pitch range (the onset is produced at 407 Hz) and has an upstep in pitch of fifteen semitones, relative to the offset of the preceding unit. Note that the measurements regarding the upsteps in pitch are taken from the offset of *yep* and the onset of *okay* not the highest point at the end of the first syllable of *okay*. These measurements can therefore often be rather conservative rather than generous.

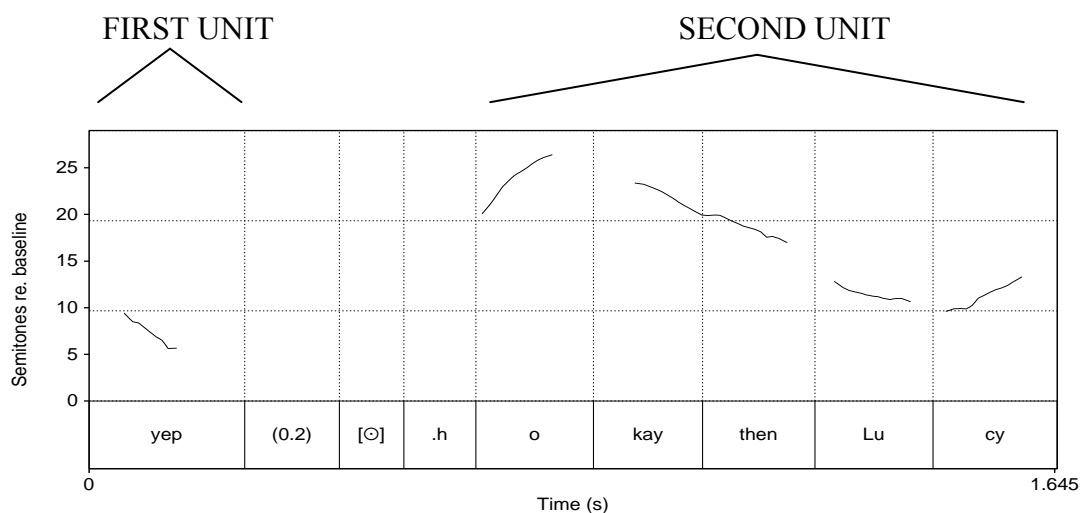


Figure 3.7: Pitch trace to show upstep in onset of second unit of F1 turn relative to offset of first (taken from Holt.C.85.4/politics/)

The upstep in pitch in the onset of the second unit, relative to the offset in pitch of the preceding unit is a consistent feature across all of the F1 turns. Figure 3.8 provides a graph which shows the upstep of the second unit in all the 24 F1 turns in the collection, as a percentage of each speaker's pitch range. Observe that although the upstep can range from 7% to 53% of a speaker's range, there are no instances in which the pitch of the second unit is not produced with an upstep in pitch relative to the first.

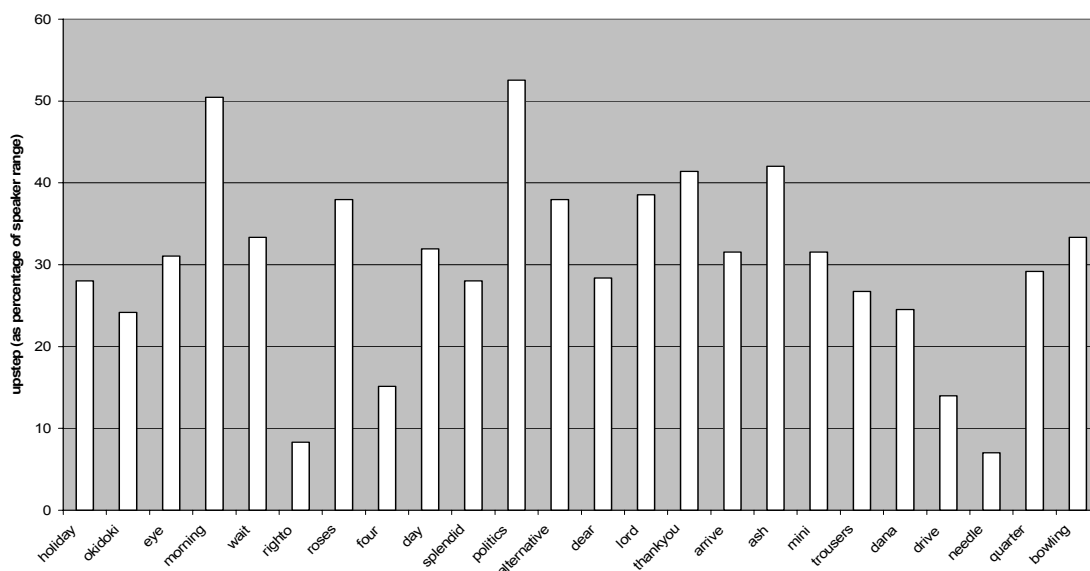


Figure 3.8: Graph to show upstep in onset of second unit relative to offset of first in all F1 turns (upsteps are calculated in semitones then expressed as a percentage of each speaker's individual pitch range)

That the second units of the F1 turns are typically initiated with an upstep in pitch and that this upstep typically results in the onset of the second unit being placed relatively high in the speakers' pitch range provides support for the claims in the literature that speakers mark out new sequences in talk-in-interaction with high onsets (Couper-Kuhlen 2001a, 2001b, 2003, 2004b; Wichmann 2000). This is because the upsteps in pitch in the second units of the F1 turns initiate the new sequences undertaken in the unit; that of call closure.

3.4.6 Glottalisation in Onset of Second Unit

A further recurrent phonetic property which routinely occurs in the onset of the first syllable of the second unit in the F1 turns is the production of either a glottal stop, creaky voice or both, all of which will be referred to under the general term of 'glottalisation'. Note that this unit-initial glottalisation only occurs on lexical items which are initiated with a vocalic portion, such as *okay* and *alright* (and not on other items such as *right* and *righto*). However, as *okay* comprises the most frequent lexical item in initial position in the second unit of the F1 turns, glottalisation is frequently found in this initial location. Glottalisation is found in the initial portion of the second unit in many of the F1 turns in the above fragments, as can be seen in their corresponding transcriptions given below¹⁰:

- Fragment 1: [jɪ̃p̃ʔ̃ (0.2) ⊙ (.) ʔəʊ:kʰei] (L4/20)
 Fragment 2: [jĩp̃ʔ̃ (0.2) ⊙ ʔʌkʰeʔ̃] (L11)
 Fragment 3: [jɪ̃p̃ʔ̃ (0.2) ⊙ .h ʔəɪkʰei] (L10)
 Fragment 4: [həhəʔ̃ m̃ (0.5) ⊙ ! .hhhh ʔɔ:kɛi] (L10)
 Fragment 5: [jĩɛs: ʔəkʰə] (L4)
 Fragment 7: [ʔm̃:: (0.5) ⊙ ! .hhhʔ̃ (0.2) ʔəʊkʰe] (L12)
 Fragment 8: [θɪ̃nsə ʔəkʰe] (L9)

¹⁰ No phonetic transcription has been given for fragment 6, as the recording is bad quality. However, it does appear to also be produced with glottalisation in the onset of the second unit of the F1 turn.

The typical occurrence of glottalisation in the onset of the second unit of the F1 turns supports the claims in the literature (see for e.g. Dilley et al. 1996) that glottalisation is found at phrasal boundaries in talk, as the onset of *okay* initiates the onset of a new unit in the F1 turn. However, these findings also suggest that not only can glottalisation be a marker of phrasal boundaries but it can also index the onset of a new *action* sequence in talk, as in F1 turns, glottalisation routinely initiates the onset of the second unit in which, unlike the preceding unit, which functions to close down the preceding talk, the new action of call closure is offered and made relevant.

3.4.7 Differences in Amplitude between Units

A further regularity in the phonetic characteristics of F1 turns is found in the relative amplitudes of the two units. Typically, the second unit is noticeably louder than the first unit. In fragments 2 to 4, and 7 to 9 above, the first syllable of the second unit can be heard and instrumentally measured as being louder than the final syllable of the first unit. Some measurements of the increases in amplitude in the onset of the second units in the above F1 turns are provided in table 3.3 below.

Fragment Number	Number of dB increase in onset of second unit
2	6 dB
3	4 dB
4	3 dB
7	2 dB
8	2 dB
9	4 dB

Table 3.3: Increase in loudness in onset of second unit of F1 turn relative to first

Figure 3.9 provides a wave form and an intensity trace of an F1 turn in which the upstep in loudness in the second unit of the turn, relative to the first, can be clearly seen in the intensity trace (this is produced by Ilene and is not discussed in the above fragments). Observe that the first unit, which is comprised of *yep*, has a peak intensity measure of around 60 dB. Then, after a 0.2 second gap and a long inbreath, the second unit is

initiated with *alright*, which is noticeably louder than the preceding *yep*, and which has a measurable upstep in intensity of 6dB.

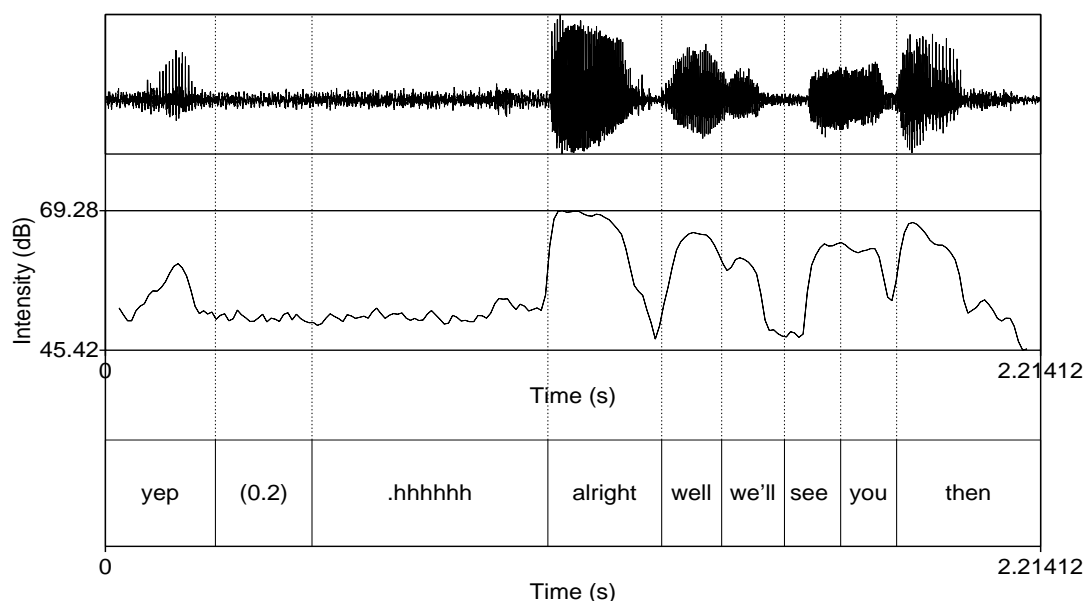


Figure 3.9: Wave form and intensity trace to show upstep in intensity of onset of second unit of F1 turn relative to offset of first unit (taken from Heritage II.2.3/thankyou/)

The upstep in loudness in the second unit of F1 turns functions alongside the other phonetic parameters discussed above, such as the pitch and the glottalisation, to index the onset of a new sequence of talk. This interrelationship of the parameters of pitch and loudness in serving to indicate a new action in talk corroborates previous claims in the literature that speakers can draw on these and a whole host of other phonetic parameters when signalling new actions and other interactional activities in talk-in-interaction (see e.g. French & Local 1983; Local et al. 1985; Local et al. 1986; Local & Kelly 1986; Local 1992, 1996, 2004; Auer 1996; Wells & Peppé 1996; Wright 2001, 2004; Ogden 2001, 2003, 2004; Ogden et al. 2004; Curl 2002, 2004; Local & Walker 2005; Walker 2003, 2004; Szczepiek Reed 2004; Couper-Kuhlen 2004).

3.4.8 Inter-Unit Clicks

The final phonetic regularity in the F1 collection is the regular occurrence of an audible alveolar or bilabial click between the first and second units of the F1 turns. Unit-medial clicks are produced in the turns in fragments 1 to 4, 7 and 9 above, which are repeated below (in phonetics) for ease of reference. Observe that the distribution of clicks is systematic: they are always produced after a unit-medial gap of either 0.2 or 0.5 seconds. Moreover, the frontwards closure of the click often shares its place of articulation with the offset of the first unit (see fragments 1-3 and 9); although this is not always the case - see fragments 4 and 7 where the offset of the first unit is produced with labiality but the click release is with alveolarity (yet it does have an initial percussive bilabial release (Pike 1943)). Note also that the production of the click can occur either with the simultaneous initiation of an inbreath (see fragments 3, 4, 7 and 9,) or without (see fragments 1 and 2).

- Fragment 1: [jɪ̃p̃ʰ (0.2) ɔ̃ (.) ʔəu:kʰei] (L4/20)
 Fragment 2: [jĩp̃ʰ (0.2) ɔ̃ ʔɹkʰẽ] (L11)
 Fragment 3: [jɪ̃p̃ʰ (0.2) ɔ̃.h ʔɹkʰẽ] (L10)¹¹
 Fragment 4: [həhəʔ m̃ (0.5) ɔ̃! .hhhhh ʔɹkei] (L10)
 Fragment 7: [ʔm̃:: (0.5) ɔ̃! .hhhʔ (0.2) ʔəukʰẽ] (L12)
 Fragment 9: [gud̃ʔ (0.5) !.hhh ɹaɪtʰəu] (L6)

Figures 3.6 and 3.10 (the former of which is repeated below) provide a wave form and a spectrogram of an F1 turn in which the speaker produces a unit-medial bilabial click. Note that 3.10 is an expanded version of 3.6 in which the environment surrounding the

¹¹ This click is different from the others: it has a low amplitude and no posterior closure. It is, however, regarded as being a click due to it having ingressive airflow and occurring at the same location in sequences as the other clicks. Further discussion on the different types of clicks that can occur in this sequential location is provided in 4.2 below.

click can be seen in more detail. The first unit of the F1 turn in these figures is terminated with bilabial closure ([jⁱɛs:p^ɪ]), which is maintained through the subsequent 0.2 second gap. This closure is then released in the onset of the following bilabial click, which can be seen in the wave form and spectrogram as a sudden high energy spike. Moreover, notice (perhaps more clearly in the wave form) that there are two spikes of energy in close proximity which coincide with the production of the click, as all clicks have two points of occlusion in the vocal tract – an anterior and a posterior closure (Beach 1938; Ladefoged & Maddieson 1996): the first of these spikes corresponds to the release of the anterior, in this instance bilabial, closure and the second spike to the posterior closure, which occurs with velarity. Notice also that the click is produced in the initiation of a following long and loud inbreath after which, the second unit, comprising of *okay then*, is produced.

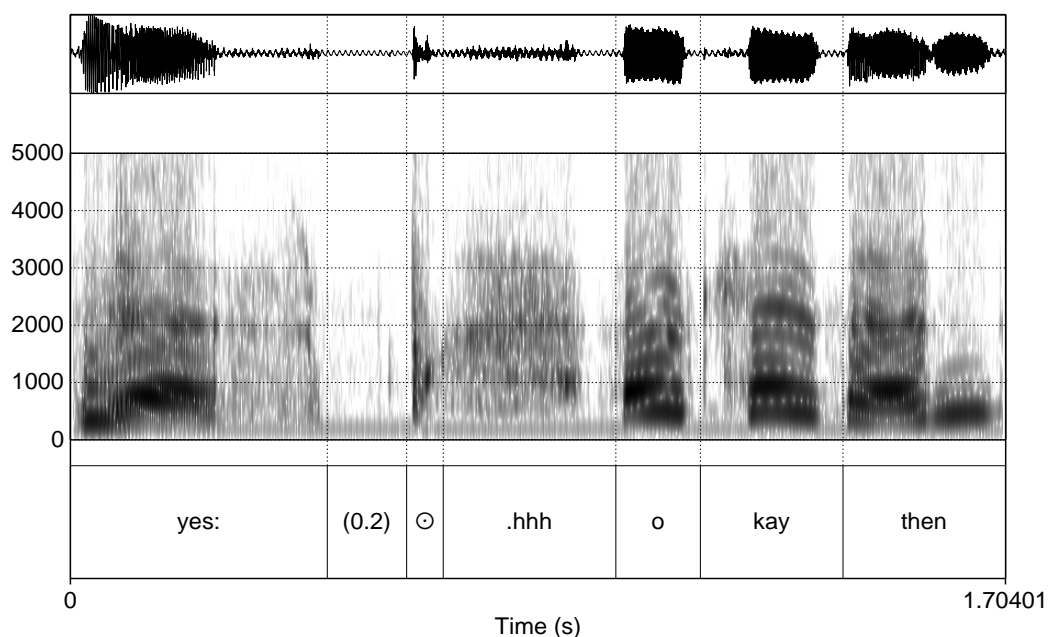


Figure 3.6: (repeated) to show the inter-unit bilabial click in the F1 turn (taken from Holt.SO.88.1.10/ash/)

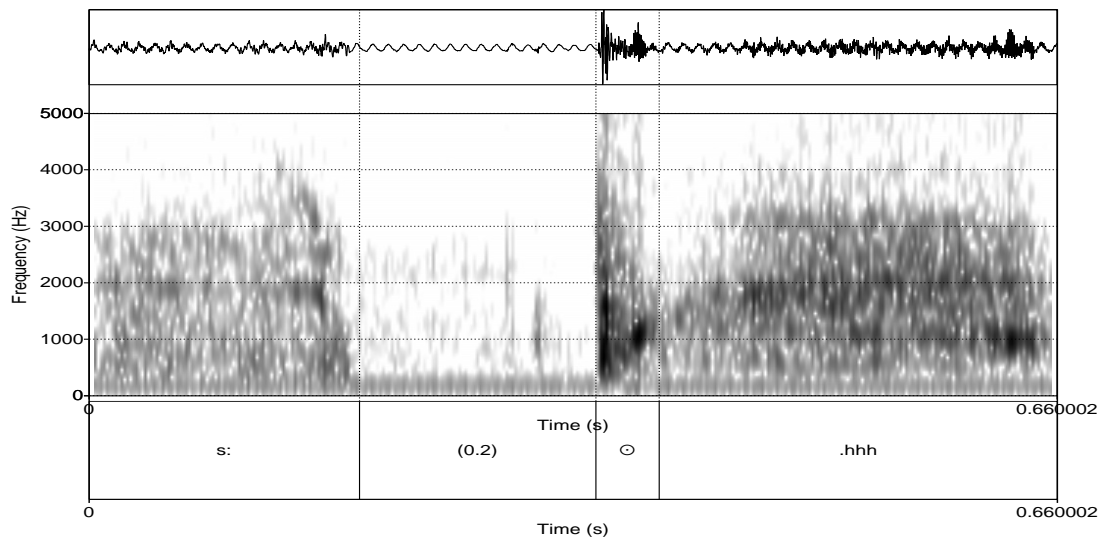


Figure 3.10: Expanded wave form and spectrogram to show inter-unit bilabial click (taken from Holt.S0.88.1.10/ash/)

The systematic distribution of audible clicks between the first and second units of the F1 turns is remarkable and their routine occurrence in this sequential location suggests that they are more than merely an articulatory consequence of a movement out of a previously maintained closure. There are two pieces of evidence to support this suggestion: (1) not all closures in talk are marked by a click in their release; and (2) the clicks do not always share the same place of articulation with the preceding closure (suggesting that they are not merely an articulatory consequence of the speaker moving their vocal apparatus). The latter claim is further supported by the finding that speakers can produce, for example, a stretch comprising an initial bilabial percussive (which is the result of the mouth being opened) and then a following alveolar click (see fragments 4 and 7). The claim that the clicks in these F1 turns are indexing a paralinguistic meaning, such as irritation or annoyance, as is typically argued to be the function of clicks in English conversation, is also disregarded, as no evidence can be found of the speakers displaying such emotive states in these F1 turns (see e.g. Abercrombie 1967; Gimson 1970; Crystal 1987; Laver 1994; Ladefoged 1982, 2001). Instead, it appears that the distribution of clicks unit-medially in F1 turns is interlinked with the interactional structure of the turn; that is, these clicks serve to further index the

demarcation between the two units in which the two distinct and disjointed actions of closing down one sequence of talk and of initiating another are undertaken. This suggestion is, to the best of the author's knowledge, the first time that clicks have been argued as being bound up with the interactional structure of naturally-occurring conversation.

3.5 Summary and Discussion

This chapter has presented an analysis of the phonetics-interaction interface which began from the collection of interactionally identical items: multi-unit F1 turns in call closing sequences. It has examined the sequential, interactional and phonetic properties of these turns and has shown that they are one device which speakers regularly employ in telephone interaction in order to shift the direction of the conversation from some on-topic talk into the closing sequence of the call. This shift is managed by the F1 turns having a (minimal) two-unit design in which two distinct actions are undertaken: the first unit attends to the preceding sequence and serves to close it down whereas the second unit initiates, offers and makes relevant the subsequent action of call closure.

In addition to regularities in their sequential and interactional designs, the F1 turns in the collection are also produced with highly systematic phonetic patterns. The *first units* in the F1 turns are typically produced with:

- a lower pitch than the following unit, which is also low in the speakers' range;
- a narrow pitch span relative to the following unit;
- a non-rising unit-final pitch contour, i.e. the pitch is either flat or falling;
- complete closure at the end of the final syllable of the unit;
- a lower amplitude than the following unit.

In comparison, the *second units* in the F1 turns are generally produced with:

- a higher pitch than the preceding unit, placed relatively high in the speakers' range;
- a wide pitch span relative to the preceding unit;
- a non-falling unit-final pitch contour, i.e. the pitch is either flat or rising;
- a marked upstep in pitch in the onset of the first syllable of the second unit of the turn relative to the offset of the pitch of the preceding unit;
- glottalisation in the onset of the second unit;
- a higher amplitude than the preceding unit;
- an inter-unit audible click.

These varied bundles of phonetic parameters with their numerous patterns distributed at different sequential locations in the F1 turns function alongside the lexis and sequential organisation to index the interactional structure of the talk. The closure at the end of the first unit, for example, along with the non-rising pitch contour and the narrow pitch span collectively perform the task of marking out the preceding sequence as being closed down and the maintenance of the closure indexes that the speaker has more talk to produce. Similarly, the upstep in pitch in the onset of the second unit of the F1 turn, alongside the glottalisation, increases in loudness and global adjustments in the organisation of pitch, all serve to indicate that the second unit undertakes a new and disjunctive sequence from that in the preceding unit. Perhaps most interesting of all, however, is the common occurrence of a click unit-medially between the two units. The fact that these clicks systematically distribute between the closing down of one sequence and the initiation of a new sequence suggests that they are not accidental. Instead, they appear to be a resource which speakers can use in order to indicate and demarcate sequence boundaries in talk. The next chapter (4) will examine this suggestion—that speakers can use clicks as an interactional resource to indicate the sequential structure of their talk-in-interaction—further.

To conclude: the F1 turns in the collection demonstrate complex, systematic and recurrent mappings between their phonetic properties, their interactional structure, their lexis and their location in sequence. These findings therefore differ markedly from the claims in the literature that first-closing turns in talk-in-interaction are typically produced by lexical items such as *well*, *okay* and *so* which stand alone in their own turn and are produced with a downwards intonation contour.