
Quantifying Mutual-Understanding in Dialogue

Marcus Colman

Submitted for the degree of Doctor of Philosophy

Queen Mary University of London

2012

Declaration

I declare that the work presented in this thesis is my own work carried out under normal terms of supervision and that the research reported here has been carried out by myself unless otherwise indicated.

Quantifying Mutual-Understanding in Dialogue

Marcus Colman

Abstract

There are two components of communication that provide a natural index of mutual-understanding in dialogue. The first is Repair; the ways in which people detect and deal with problems with understanding. The second is Ellipsis/Anaphora; the use of expressions that depend directly on the accessibility of the local context for their interpretation. This thesis explores the use of these two phenomena in systematic comparative analyses of human-human dialogue under different task and media conditions. In order to do this it is necessary to a) develop reliable, valid protocols for coding the different Repair and Ellipsis/Anaphora phenomena b) establish their baseline patterns of distribution in conversation and c) model their basic statistical inter-relationships and their predictive value. Two new protocols for coding Repair and Ellipsis/Anaphora phenomena are presented and applied to two dialogue corpora, one of ordinary 'everyday' conversations and one of task-oriented dialogues. These data illustrate that there are significant differences in how understanding is created and negotiated across conditions. Repair is shown to be a ubiquitous feature in all dialogue. The goals of the speaker directly affect the type of Repair used. Giving instructions leads to a higher rate of self-editing; following instructions increases corrections and requests for clarification. Medium and familiarity also influence Repair; when eye contact is not possible there are a greater number of repeats and clarifications. Anaphora are used less frequently in task-oriented dialogue whereas types of Ellipsis increase. The use of Elliptical phrases that check, confirm or acknowledge is higher when there is no eye contact. Familiar pairs use more elliptical expressions, especially endophora and elliptical questions. Following instructions leads to greater use of elliptical (non-sentential) phrases. Medium, task and social norms all have a measureable effect on the components of dialogue that underpin mutual-understanding.

It is probable that before the dawn of the twentieth century this prophetic picture will have been surpassed in actual fact, and the telephone will be a quite indispensable element in English social life. But it will be a much more comprehensive and effective instrument than the telephone as we know it at present, and the likelihood is that it will be fitted in our houses just as gas or electricity is now.

(...)

It will make millions merry who have never been merry before, and will democratize, if we may so write, many of the social luxuries of the rich.

(...)

There is, indeed, no element in our social life which will be unprovided for, and if, as it is said to be not unlikely in the near future--the principle of sight is applied to the telephone as well as that of sound, earth will be in truth a paradise, and distance will lose its enchantment by being abolished altogether.

(...)

The Pleasure Telephone opens out a vista of infinite charm which few prophets of to-day have dreamed of, and who dare to say that in twenty years the electric miracle will not bring all the corners of the earth to our own fireside?

Arthur Mee
'The Pleasure Telephone'.
The Strand, 1898.

In physical science the first essential step in the direction of learning any subject is to find principles of numerical reckoning and practicable methods for measuring some quality connected with it. I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of Science, whatever the matter may be.

Baron William Thomson Kelvin
'Electrical Units of Measurement'.
Popular Lectures and Addresses, Vol. 1, 1889.

Table of contents

Acknowledgements	8
List of tables	9
List of figures	12
Chapter 1: Introduction	13
1.0 Background	13
1.1 Scope of the thesis	14
1.2 Research questions	15
1.3 Contributions	16
1.4 Overview of the thesis	17
1.5 Review	18
Chapter 2: Mutual understanding in dialogue	19
2.0 Introduction	19
2.1 Factors affecting mediated communication	21
2.1.1 Independent and dependent variables in previous studies	22
2.1.2 Comparing mediated to face to face communication	23
2.2 Theories of communication media	29
2.2.1 The grounding/collaborative model	29
2.2.2 Media Richness Theory	34
2.2.3 Dialogue coding approaches	38
2.3 Quantification of mutual understanding in dialogue	44
2.3.1 Accessing elided material in dialogue	45
2.3.2 Ellipsis: shared context in dialogue	48
2.3.3 Shared context in dialogue: anaphora	49
2.3.4 Shared context in dialogue: non-sentential utterances	50
2.3.5 Shared context in dialogue: sentential ellipsis	52
2.3.6 Ellipsis and levels of participation	52
2.4 Conversation Analysis	54
2.4.1 Repair and repair initiations in dialogue	55
2.4.2 The validity of applying the CA model	56
2.4.3 Next turn repair initiations (NTRI)	60
2.4.4 A special case of NTRI: Non-Component Referents	61
2.4.5 Disfluencies and self-repairs	62
2.4.6 Continuations and third turn repair	63
2.5 Quantifying mutual understanding	64
2.6 Review	66
Chapter 3: Methodology	67
3.0 Overview	67
3.1 General methods	67
3.1.1 Methodology: experimental corpora	67
3.1.2 The British National Corpus	67

3.1.3 The HCRC map task.....	68
3.2 Statistical methods.....	70
3.2.1 Methodology: IV and DV.....	70
3.2.2 Methodology: reliability of coding schemes.....	71
3.2.3 Methodology: inferential statistical tests.....	73
3.2.4 Methodology: nonparametric statistical procedures.....	73
3.2.5 Methodology: statistical correlations.....	74
3.2.6 Methodology: raw and adjusted data.....	74
3.3 Positive indices of mutual understanding: presumed shared context.....	75
3.3.1 Glossary of terms.....	75
3.3.2 The ‘ellipsis’ coding protocol.....	77
3.3.3 Application of the presumed shared context coding scheme.....	79
3.3.4 Reliability of the ‘indices of presumed shared context’ coding scheme.....	80
3.4 Negative indices of mutual understanding: repair and repair initiations.....	82
3.4.1 Glossary of terms.....	82
3.4.2 Construction of the repair protocol.....	84
3.4.3 Application of the repair protocol.....	85
3.4.4 Repair protocol reliability.....	86
3.4.5 Uncoded CA repair phenomena.....	87
3.4.6 Repair or disfluency?.....	88
3.5 Review.....	90
Chapter 4: Positive indices of mutual understanding in the HCRC map task corpus	92
4.0 Overview.....	92
4.1 Method.....	92
4.2 Results.....	92
4.2.1 Descriptives.....	92
4.2.2 Overall distribution of indices of presumed shared context.....	94
4.2.3 Medium and familiarity differences.....	95
4.2.4 Role differences.....	97
4.3 Correlational data.....	99
4.3.1 Correlations between coding categories.....	99
4.3.2 Accuracy correlations.....	100
4.4 Summary of results.....	100
4.5 Discussion.....	101
4.6 Review.....	102
Chapter 5: Negative indices of mutual understanding in the HCRC map task corpus	104
5.0 Overview.....	104
5.1 Method.....	104
5.2 Results.....	104
5.2.1 British National Corpus vs. HCRC maptask data.....	104
5.2.2 Repair Frequencies.....	106
5.2.3 Medium, familiarity and role differences.....	107
5.2.4 Specificity of next-turn repair initiations.....	110
5.3 Relationships between variables.....	112
5.4 Summary of results.....	114
5.5 Discussion.....	114
5.6 Review.....	116
Chapter 6: The relationship between positive and negative indices of mutual understanding.....	117
6.0 Overview.....	117

6.1 Comparison of results from all approaches	118
6.1.1 Accuracy and instructions	119
6.1.2 Rejections and continuations	120
6.1.3 Formulation repairs	121
6.1.4 Polar answers associated with accuracy	122
6.1.5 Grounding and acknowledgements.....	122
6.1.6 Further evidence for grounding	124
6.1.7 Correlation and regression	124
6.2 Multiple Regression Analysis.....	124
6.2.1 Further regression tests	126
6.2.2 NTRIs and NSU questions	129
6.2.3 Summary of results.....	131
6.3 Preliminary conclusions	131
6.3.1 Review of correlational results	132
6.4 Factors affecting successful interaction.....	132
6.4.1 Further dimensions	133
6.5 Discussion	134
6.6 Review.....	137
Chapter 7: Conclusions	138
7.0 Recap of thesis.....	138
7.1 Statistical results.....	139
7.2 Models of interaction.....	141
7.3 Implications for models of interaction.....	145
7.4 Further applications of the methodology	147
7.5 Conclusions and further work.....	150
7.6 Final thoughts	152
References	153
Appendix A: Glossary of dependent measures terms and abbreviations	172
Appendix B: Presumed shared context exemplars from the transcripts.....	175
Appendix C: Repair and repair initiation exemplars from the transcripts.....	179
Appendix D: Confusion matrices	183
Appendix E: Medium and familiarity repair and shared referents (ellipsis) distributions	185
Appendix F: Correlations of ‘ellipsis’ categories.....	191

Acknowledgements

The Human Communication Research Centre (HCRC) group at the Universities of Glasgow and Edinburgh, Scotland, UK created, collated and made available the map task data. Amy Isard and Jean Carletta from the HCRC group were especially helpful in ensuring that I had all of the data that I needed. Bethan Davies filled me in on some things I didn't know. Greg Mills provided plenty of moral support and beer when I was thirsty. Thanks Jim. Jackie Folkes supported me in every other way, thanks honey. My friend Aisa gave me lots of helpful advice, probably more than she realises. Professor Pat Healey supervised this thesis. I quite literally do not know how to thank him for everything. Thanks Pat, I owe you.

This thesis is dedicated to my son Zachary.

List of tables

Table 1: examples of media in terms of audio-visual mode and synchronicity.....	21
Table 2: task-richness requirements according to Media Richness Theory (adapted from Hollingshead et al., 1993).....	35
Table 3: characteristics of different modes of communication according to media synchronicity theory (adapted from Dennis and Valacich, 1999).....	37
Table 4: ‘moves’ from Dialogue Games Analysis (Carletta et al., 1996).....	39
Table 5: example taxonomy of non-sentential utterances (NSUs) taken from Fernandez, 2006	51
Table 6: Non-Sentential Utterance taxonomy used in this thesis, taken from Colman, Eshghi and Healey (2008).....	51
Table 7: dependent variables: categories of dialogue phenomena to be quantified and used in the analysis.....	71
Table 8.1: glossary of positive indices of mutual understanding: indices of presumed shared context.....	76
Table 8.2: glossary of positive indices of mutual understanding (continued)	77
Table 9: indices of ‘presumed shared context’ coding protocol confusion matrix: ...	80
Table 10: categories of repair and repair initiation identified by the coding protocol	83
Table 11: indices of ‘repair and repair initiation’ confusion matrix	86
Table 12: position one ‘repair’ categories correlated with ‘disfluency’ codings from the HCRC map task.....	89
Table 13: exact figures for BNC and HCRC map task ‘indices of presumed shared context’ coding.....	93
Table 14: medium and familiarity differences for the ‘presumed shared reference’ codings in the HCRC map task (ANOVA).....	95

Table 15: ‘role’ differences in the ‘presumed shared reference’ coding of the HCRC map task (ANOVA)	98
Table 16: rates of ‘anaphora’ and ‘NSUs’ by ‘task role’ in the HCRC map task.....	99
Table 17: significant correlations of ‘presumed shared referent’ categories with deviation from the route (accuracy) in the HCRC map task	100
Table 18: relative frequency of ‘repair’ categories in the HCRC map task and a BNC sample	105
Table 19: ‘repair’ differences in audio versus face-to-face conditions in the HCRC map task using standardised data (ANOVA).....	107
Table 20: ‘repair’ differences in familiar versus unfamiliar conditions in the HCRC map task using standardised data (ANOVA).....	107
Table 21: ‘P2NTRI specificity level’ differences in familiar versus unfamiliar conditions in the HCRC map task (Mann-Whitney tests).....	108
Table 22: ‘repair’ category differences in ‘role’ conditions in the HCRC map task (ANOVA).....	110
Table 23: ‘P2NTRI’ repair category correlations within the ‘follower’ role in the HCRC map task	112
Table 24: ‘repair’ categories correlating with different ‘follower NTRI’ specificity levels (1=lowest; 5=highest) in the HCRC map task.....	113
Table 25: correlational relationships between ‘repair’ categories in both ‘role’ conditions in the HCRC map task.....	113
Table 26: categories from ‘repair’, ‘presumed shared referents’ and ‘DGA’ coding schemes that significantly correlate with ‘deviation from the route’ using standardised scores.....	119
Table 27: stepwise multiple regression of predictors of ‘deviation from the route’ using raw frequencies from ‘repair’ and ‘presumed shared reference’ categories (only significant predictors are included).....	125
Table 28: stepwise multiple regression of predictors of ‘deviation from the route’ using standardised data from ‘repair’, ‘presumed shared reference’ and ‘DGA’ categories (only significant predictors are included)	125

Table 29: stepwise multiple regression of predictors of ‘high specificity NTRIs’ from ‘repair’ and ‘presumed shared reference’ categories (only significant predictors are included).....	127
Table 30: stepwise multiple regression of predictors of ‘presumed shared context’ using ‘repair’ categories (only significant predictors are included).....	128
Table 31: stepwise multiple regression of predictors of ‘repair’ categories’ from ‘presumed shared reference’ categories (only significant predictors are included).	128
Table 32: stepwise multiple regression of predictors of ‘anaphora’ from ‘repair’ categories (only significant predictors are included)	129
Table 33: significant correlations between ‘giver NTRI’ specificity levels 1-5 and ‘presumed shared context’ categories	130
Table 34: significant correlations between ‘follower NTRI’ specificity levels 1-5 and ‘presumed shared context’ categories	130

List of figures

Figure 1: giver and follower maps	69
Figure 2: indices of presumed shared context/‘ellipsis’ coding protocol (anaphora and answers).....	78
Figure 3: indices of presumed shared context/‘ellipsis’ coding protocol (questions and statements).....	79
Figure 4: the ‘repair and repair initiation’ coding protocol.....	85
Figure 5: frequency of ‘presumed shared context’/‘ellipsis’ categories in a sample of dialogue from the BNC	94
Figure 6: frequency of ‘presumed shared context’/‘ellipsis’ categories in the HCRC map task corpus.....	94
Figure 7: frequency of ‘presumed shared context’/‘ellipsis’ categories in the ‘giver’ role in the HCRC map task corpus.....	97
Figure 8: frequency of ‘presumed shared context’/‘ellipsis’ categories in the ‘follower’ role in the HCRC map task corpus	97
Figure 9: frequency of ‘repair and repair initiation’ categories in a sample from the BNC	106
Figure 10: frequency of ‘repair and repair initiation’ categories in the HCRC map task corpus.....	106
Figure 11: frequency of ‘repair and repair initiation’ categories in the ‘giver’ role in the HCRC map task corpus	108
Figure 12: frequency of ‘repair and repair initiation’ categories in the ‘follower’ role in the HCRC map task corpus.....	109
Figure 13: frequency of ‘next-turn repair initiation’ specificity levels 1-5 in the ‘giver’ role in the HCRC map task corpus.....	111
Figure 14: frequency of ‘next-turn repair initiation’ specificity levels 1-5 in the ‘follower’ role in the HCRC map task corpus	111

Chapter 1

Introduction

1.0 Background

This introductory chapter will outline the background to this thesis; its scope; the research questions examined and provide an overview of the overall thesis.

This thesis deals with the quantification of elements of communication; specifically, reliable quantitative evidence for how a shared context is achieved, understood and used by all parties involved in communication. From a broad perspective this could be seen as part of a wider search for experimental measures that are less susceptible to criticisms of a lack of ecological validity. This has been a fundamental problem in disciplines such as experimental social psychology, which may be reduced to asking participants to self-report behaviour or attitudes using Likert scales or similar with all attendant problems such as social desirability or demand characteristics (Orne, 1962).

Similar problems have been found in studies evaluating computer-mediated-communication (CMC); see e.g. Whittaker (2002); Lazar et al., (2010). Studies of mediated communication have reported unclear or contradictory findings regarding the relative benefits of different media using dependent measures such as time taken over tasks, self-reported satisfaction, or measures such as number of turns or words. These measures do not identify the core reasons that may make one medium better than another for a given task. This thesis examines the hypothesis that the structure of interaction is a function of many variables including medium, social norms and goals. The analyses presented in this thesis demonstrate a method for comparing and contrasting dialogue. This method is based upon the transparent, observable and

measurable features of communication that are critical to mutual-understanding. References to pronouns such as ‘I like *that*’, if they are not questioned, and do not lead to later trouble, demonstrate that the referent object or antecedent of the pronoun is understood. This also applies to elliptical or non-sentential utterances which require accessing a shared context in order to be fully understood; this is argued to positively index mutual-understanding. If there is a request for clarification, or trouble emerges due to the unclear use of anaphora or elliptical phrases (e.g. ‘What?’ or ‘Who do you mean?’) then a repair sequence begins (see Schegloff et al., 1977). This provides an index of both trouble in mutual-understanding and how easily it is dealt with.

This thesis describes an original approach to examining the creation of mutual understanding in communication based on observable instances of the creation and negotiation of shared context. Here I specifically focus on *mutual-understanding*, the process by which speakers ensure that they share the content and meaning of the salient topics and references. This thesis argues that mutual-understanding of context is achieved through the process of repair (e.g. Schegloff, 2010) and demonstrated through the use of shared context, also referred to as ‘ellipsis’¹ (e.g. Eshghi and Healey, 2009) and that these processes can be identified and measured in a way that is valid and reliable.

1.1 Scope of the thesis

This thesis will examine the issue of whether reliable quantifiable measures of mutual understanding can be established, and, as applied to the HCRC² map task corpus, be used to examine the communicative process in the medium conditions of face to face dialogue versus speaking with no possible eye contact. Further investigations appropriate to the data also examine the effects and relationship of

¹ In this thesis anaphora, non-sentential utterances and sentential ellipsis are together referred to as ‘presumed shared context’; referred to as ‘ellipsis’ in Colman, Eshghi and Healey (2008).

² The Human Communication Research Centre, Universities of Edinburgh and Glasgow, UK.

these with regard to task type and normative behaviour associated with familiarity. The findings may inform models of interaction in task-based or goal-oriented dialogue. The distribution of these phenomena are presented and discussed.

In this thesis there are certain terms that will be dispreferred such as ‘breakdowns’ or ‘miscommunication’^{3 4} despite their common occurrence in research on dialogue as these terms may be misconstrued. A true ‘breakdown’ would involve a power cut, dropping a ‘phone or similar (discussed in Scrivener et al., 1996); the trouble source - repair sequence is as much a natural part of dialogue as anything else and nothing within that is ‘broken’. ‘Miscommunication’ may be understood to be e.g. when two people leave a conversation each thinking that their next meeting is on a different day; again, this problem of ‘correctness’ is not being investigated.

It is important to note that this thesis is not a study of formal linguistics or computational linguistics, although there may be parts of this thesis which could potentially inform the latter as in the work of Purver (2004) and Fernandez (2006) which looked at clarification requests and non-sentential utterances respectively. Conversely, this thesis *is* a methodology for linguistically naive researchers to examine any dialogue transcripts and identify instances of anaphora, non-sentential utterances, sentential ellipsis, repair and repair initiations; which are argued to index both positive and negative aspects of mutual understanding.

1.2 Research questions

The questions to be addressed by this thesis are:

- Can categories of anaphora, non-sentential utterances and sentential ellipsis (referred to here as ‘indices of presumed shared context’) be reliably identified?
- Do measures of ‘presumed shared context’ discriminate between face-to-face/audio media; familiarity; task type?

³ Except when referring to other authors’ work where these terms have been used.

⁴ Preferred terms are ‘trouble’; ‘trouble source’; repair sequence’; ‘repair’.

- Can categories of repair and repair initiations be reliably identified?
- Do categories of presumed shared context or repair and repair initiations differ between naturally-occurring and task-oriented dialogue?
- Do categories of presumed shared context or repair and repair initiations predict success in the HCRC map task?
- Are there significant relationships between positive indices of mutual understanding (measured by categories of presumed shared context) and negative indices of mutual understanding (measured by categories of repair and repair initiations)?

1.3 Contributions

The Map Task corpus of task-oriented dialogues was designed and created by the Human Communication Research Centre at the universities of Edinburgh and Glasgow, UK. The British National Corpus of naturally-occurring spoken dialogue is part of a larger corpus collected by the BNC Consortium.

Work published by Healey (1999) and Healey and Thirlwell (2002) first introduced the idea of identifying types of repair as a quantitative measure of coordination in order to contrast different modes of communication. The present author engaged in a series of discussions with P. Healey and M. Thirlwell in an attempt to refine and improve the coding scheme; the present author designed a method of demonstrating validity and reliability which was published in Healey, Colman and Thirlwell (2005). The repair coding protocol had by this time gone through several iterations; this (2005) version was thought to be sufficiently robust to begin coding the HCRC map task dialogues (carried out solely by the present author by annotating by hand printouts of each line of dialogue for all 128 map tasks). As the coding protocol was applied to more dialogues, it became apparent that two of the categories were so rare as to be unnecessary; this is reflected in the version of the coding protocol presented in this thesis. The layout and illustration of this version of the repair protocol was created by the author. Discussions between P. Healey, A. Eshghi and the author led to the coding protocol for ellipsis and anaphoric reference; ellipsis had been examined previously in Eshghi and Healey (2009); Healey et al. (2008) but without a

standardised reliable coding protocol. A similar process to the development of the repair protocol took place; intensive discussions led to over ten iterations of the protocol, each of which were critiqued and applied to examples (by Eshghi and the author) until the current coding protocol was agreed upon (the layout and illustration of the protocol was created by the author and previously published in Colman et al., 2008). The author repeated the process of hand annotating dialogue transcripts of all 128 map tasks. The hand annotated manuscripts of the dialogues were then collated in terms of the frequencies of each coding category for each dialogue, broken down into the two task roles and this data was entered into an SPSS spreadsheet by the author. All analyses were carried out solely by the author.

The following research, some of which is presented in this thesis, was published:

P.G.T. Healey, M. Colman & M. Thirlwell (2005). "Analysing Multi-Modal Communication: Repair-Based Measures of Human Communicative Co-ordination" in *Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems*, Dordrecht: Kluwer Academic.

M. Colman, A. Eshghi and P. G. T. Healey (2008). 'Quantifying ellipsis in dialogue: an index of mutual-understanding.' In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, June 2008 Columbus, Ohio: Association for Computational Linguistics.

M. Colman and P. G. T. Healey (2011). The distribution of repair in dialogue. In *Proceedings of the 33rd Annual Cognitive Science Conference*, July 2011, Boston, Massachusetts.

1.4 Overview of the thesis

Chapter 2 describes previous experimental and theoretical work examining interaction. A methodology that is independent of task, medium or other confounds will allow previously unavailable comparisons. Some theories of how media affect interaction are explained and critiqued in favour of a conversation analysis-inspired approach to examining the content of interaction based upon the accessibility of shared context (through the use of anaphora, non-sentential utterances and sentential ellipsis) and the use of repair and repair initiations. Chapter 3 outlines the general methodology to be used. This includes a description of the corpora (BNC and HCRC map task) and statistical methods; also the use of coding protocols in capturing and

quantifying indices of mutual-understanding. Protocols for extracting indices of mutual understanding are presented along with reliability scores. Chapters 4 and 5 present results from applying the coding protocols to the BNC and HCRC map task corpora. Results cover differences across conditions and correlational data. In Chapter 6 the results from both the repair and the ellipsis codings are combined in terms of correlations. Regression analyses are used to show the predictors of accuracy and other variables in the map task. Chapter 7 contains a recap and discussion; further potential uses of this methodology are examined, including the analysis of dialogue within clinical populations.

1.5 Review

This thesis investigates the creation and negotiation of mutual-understanding of context. This will be examined through the identification and analysis of two interactional phenomena: presumed shared context (identified through the use of anaphora and ellipsis) and trouble and resolution (identified through the use of repair and repair initiations). Protocols to measure these are demonstrated and tested for reliability. These data will be used to provide evidence for how interaction is affected by medium, familiarity and task goals.

Chapter 2

Mutual understanding in dialogue

2.0 Introduction

Current fast-moving advances in communication technology have been led by the technology rather than communicative needs; it has not been clearly established how a given mode of mediated communication facilitates or impedes the process of mutual understanding. Previous studies of mediated communication have shown various – sometimes contradictory – results comparing and contrasting modes of mediated communication with face-to-face (FTF). Theories explaining empirical studies have emphasized the importance of different aspects of media and interaction. A method of investigating how mutual understanding occurs which would support quantitative statistical analysis would allow for the strengths and weaknesses of different communicative modes to be understood and measured.

Mutual understanding (also referred to as ‘intersubjectivity’, see e.g. Schegloff, 1992) is the process whereby interlocutors satisfy themselves that the intended meaning is being conveyed and understood. This thesis takes the approach that the process underlying mutual understanding is achieved through i) the construction and editing of a contribution⁵, ii) raising issues of trouble in understanding and iii) amending or rewording a contribution as a result of how it was responded to. Identifying and quantifying these processes would provide a new approach to examining how mutual understanding is created and how the process may be affected under different conditions.

Lazar et al. (2010) argue that approaches to human-computer interaction (HCI)

⁵ The term ‘contribution’ is used rather than ‘turn’ as it refers to any verbal, nonverbal, visual or other signal that incrementally adds to the dialogue ; it is neutral with regard to modality.

studies of communication are based upon methodologies taken from other disciplines such as engineering or social psychology. These methods may not be relevant to contemporary forms of technology-based human-human interaction such as videophones or messaging through social networking websites. Despite this, the methodological approaches that they condone are not taken from relatively recent human-human interaction studies that have emerged from theoretical approaches such as grounding (e.g. Convertino et al., 2008), conversation analysis (e.g. Sidnell, 2010) or dialogue games analysis (e.g. Monk, 2009); instead Lazar et al. outline experimental design, surveys, case studies and other well-established techniques.

Ginzburg (2012) makes a similar point with respect to the general issue of understanding talk or interaction;

“The major problem with this huge literature, however, is that it is compartmentalized – phoneticians ignore phonology, conversational analysts ignore discourse semantics and phonetics, students of AI ignore psycholinguistics, and so on.”
(Levelt 1993, quoted in Ginzburg 2012, p. 7)

Ginzburg argues that understanding interaction must necessarily look at issues of trouble and misunderstanding as well as shared understanding;

“The interactive stance involves taking seriously the fact that communication involves multiple agents with distinct beliefs and desires and places importance on explicating the potential for misunderstanding, rejection, and correction, as well as success.”
(Ginzburg, 2012; pp. 7-8)

Ginzburg emphasizes the need to incorporate repair and grounding into a model of interaction:

“We can suggest that the adequacy of semantic theory involves **the ability to characterize for any utterance type the contextual update that emerges in the aftermath of successful exchange and the range of possible clarification requests otherwise.**”
(Ginzburg, 2012; p. 8; Ginzburg’s emphasis)

This chapter will outline some empirical results of studies examining modes of

communication, some theories that have been put forward to explain the effects that a given medium may have upon the act of communication, and present the background to the current approach which examines mutual understanding in dialogue through the identification of positive and negative evidence for mutual understanding having taken place. This approach attempts to address the issue of how grounding and repair phenomena are achieved in dialogue.

2.1 Factors affecting mediated communication

Studies examining communication through different media have identified variables that appear to have an effect on the process and/or outcome, outlined in the following sections. Media characteristics; one of the most obvious characteristics of an interactive medium is whether it is synchronous or asynchronous; i.e. if the communication takes place in 'real-time'. Another fundamental characteristic is the sensory mode of communication; i.e. is something looked at (text or visuals), or listened/spoken to, or both?

Table 1: Examples of media in terms of audio-visual mode and synchronicity

	synchronous	asynchronous
Audio-only	Telephone	Answerphone
Visual-only	Text chat	Email
Audio-visual	Videophone/Webcams	Video messages

Although this is a simple way of categorising the characteristics of different media, by sensory mode vs. synchronicity, it is by no means the only way. Clark and Brennan (1991) suggest that there are eight characteristics that need to be taken into account for every medium; Dennis and Valacich (1999) suggest five. Table 1 illustrates a similar approach to that of Galegher and Kraut (1996); they suggest that the most relevant characteristics are interactivity (speed of response) and expressiveness (the level to which ideas can be conveyed, for example through non-verbal cues).

Research into communication and communication technologies has generally

involved laboratory studies looking at both objective (e.g. time taken; quality of task result) and subjective measures (e.g. satisfaction with the medium); comparisons are made for a given task between participants using different media. Other approaches have been taken, including interviews, survey questionnaires and ethnographic field studies. It could be argued that field studies are the most useful due to their inherent ecological validity; however, there is a trade-off for the control of variables found in lab studies. One drawback to the control allowed in experimental studies is that it may involve testing participants using a medium that they are not familiar with; in this case it may be that mediated/FTF differences would be lessened to some degree after gaining familiarity with the medium, for example in the emergence of norms regarding emotion and humour where those aspects are not easily conveyed.

2.1.1 Independent and dependent variables in previous studies

Independent variables in studies involving communication technologies generally concern manipulations of the medium, the task and the group/participant characteristics. Treating the medium as an independent variable in itself may be problematic; as it was pointed out above, different communication behaviours are supported by different media. If a difference is found in some dependent measure, it may be difficult to conclude which characteristic of the independent variable (medium) was responsible for this effect, e.g. whether asynchronicity or lack of audio was responsible for the effects found using email (Whittaker, 2002). As for tasks given in experimental studies, it does not always seem to be the case that the nature of the task is taken into account when comparing media differences. Various classifications of task have been postulated; these can be simple, as in a cognitive/problem solving vs. social/negotiation dichotomy, to more complex models such as McGrath's (1984) task circumplex. Dependent measures can be seen to fit into three general categories; outcome, process and subjective responses. Of these, outcome measures are the most widely used, for example time taken over a task using a particular medium; however, when differences are not found in outcome measures this may conceal differences in process. Process measures do not conform to a standard type; they may range from the number of words uttered to specific

examples of dialogue games (Kowtko et al., 1991). The wide range of process measures that have been devised reflects the difficulty of reducing an act of communication to its components:

"The most important observation about these data is how exceedingly difficult it is to quantify natural human communications. Perhaps the greatest source of difficulty arises from their unstructured, unruly, and error-ridden nature. When they are examined closely, natural human communications, whether oral, written, or typewritten, appear at first glance to follow few syntactic or grammatical rules."
(Chapanis et al., 1977, p122).

2.1.2 Comparing mediated to face to face communication

As mentioned above, many studies examining communication technologies compare mediated communication with FTF. It is possible that this is inappropriate when looking at new media such as internet forums (or newsgroups), or email due to the possibly extreme asynchronicity of communication, i.e. taking place over weeks or months (Whittaker et al., 2002). Models of communicative co-ordination dealing with intersubjectivity or turn-taking based upon FTF examples (e.g. Clark and Schaefer, 1989) may be irrelevant in cases of this type. Other technologies such as computer text chat can lead to problems such as high parallelism, where many participants are submitting messages simultaneously, making it very difficult to follow a conversational thread (Dennis and Valacich, 1999). In contrast, within FTF communication it is unusual for people to contribute simultaneously for longer than brief periods; there are conversational norms that are generally adhered to (McGrath, 1990; Sacks et al., 1974). However, it may be that parallelism offers advantages that are not available in all media, for example in a time-limited idea generation task (Dennis and Valacich, 1999; Dennis et al., 2008). The characteristics of a medium do not necessarily identify the behaviours that are being supported or hindered, for example paralinguistic signals. Measures of communicative efficacy have been found to be problematic; it may be necessary to examine communicative process if outcome measures do not discriminate between conditions. This may become important if it appears that a less efficient medium is demanding more effort from its

users, although leading to a similar outcome; this may be due to some difficulty in reaching mutual understanding.

An influential study in communication media research was carried out by Chapanis et al. (1972; see also Chapanis, 1975); in a series of studies task outcome measures were examined in collaborative problem solving tasks using different media combinations. The crucial medium was found to be the audio channel; in a comparison between audio only and high quality video plus audio, it was found that adding the video channel did not improve problem solving. Further comparisons demonstrated that the speech channel was crucial to collaborative problem solving; adding or removing media such as video, text, writing materials or even using FTF communication made little difference. For Chapanis this was an entirely unexpected result;

"From the voluminous literature on kinesics, gestures and 'body language' I had been led to predict a large difference between face-to-face communication and communication by voice alone. The voice channel by itself seems impoverished in comparison with the variety and richness of the information-bearing clues available in face-to-face communication. The data did not conform at all to my expectations. The average amount of time taken to solve problems by voice alone was only slightly more than it was in face-to-face communication." (Chapanis, 1975; p40)

It seems that these results were due to the particular task types used by Chapanis et al. These tasks took the form of collaborations where information was spread between two participants; for example equipment assembly, information retrieval or geographic orientation. It was claimed that tasks of this type were the most appropriate,

"Instead of being abstract or artificial puzzles of the kind often devised to measure hypothetical psychological processes, they are of recognizable and practical importance in everyday life. They have definite, recognizable solutions, which can usually be reached within approximately one hour." (Chapanis, 1975; p38).

However, in workplace situations it is only in certain instances when a task with a definite solution will be collaborated on; it is more likely that some type of negotiation will take place.

Other evidence regarding the usefulness of social cues can be found in studies of video-mediated communication. These studies should be particularly useful as they support both paralinguistic audio cues and nonverbal visual cues such as eye gaze, gestures and facial expressions, making it closest to FTF. Whittaker (1995) points out that previous work on VMC and other communication technologies has been based on the non-verbal communication hypothesis; i.e. that visual information supports the co-ordination of both process and content. This is due to the role of behaviours such as head nods in understanding (e.g. Kahneman, 1973) and facial expressions in conveying attitudes (e.g. Argyle et al., 1974). Process is influenced by non-verbal communication through behaviours such as eye gaze in managing speaker transitions or turn-taking (Argyle et al., 1968).

The role of the visual mode plays a part in all social cue theories; is there evidence that a visual channel in a medium improves the process and content of communication in the way that the non-verbal communication hypothesis would predict? However, when evaluating studies using video it should be remembered that effects such as transmission lags can disrupt communication and may confound any results (Cohen, 1982; O'Conaill et al., 1993; Whittaker and O'Conaill, 1993).

Whittaker (2002) suggests that content co-ordination processes are served by non-verbal communication in two ways; cognitive cues influence understanding through head nods and visual attention, whereas social or affective cues demonstrate attitudes or emotions through facial expressions, posture and eye gaze (Argyle et al., 1974). Process co-ordination is influenced by turn-taking cues such as head turning, posture and eye gaze (Argyle et al., 1968).

Speech would appear to be the most important factor in the co-ordination of understanding, but what about the effect of social or affective cues on co-ordination

of conversational content? It may be that affective information is not generally verbalised, but inferred; in this case it may be that a video channel would be more useful. Whether this leads to an improvement in the group's functioning will depend on the relevance of affective information; for example Anderson et al. (1997) found no improvement over audio-only for a collaborative problem solving task when two types of video channel were tested. One of these was a standard video/audio connection; the second was a 'videotunnel' that allowed direct eye contact over video through the use of mirrors.

If process co-ordination is facilitated by non-verbal turn-taking cues, then it would be predicted that process co-ordination would be more efficient through a video channel. This effect has not been found though (Sellen, 1992, 1995; O'Conaill et al., 1993; Whittaker and O'Conaill, 1993), in measures such as overlaps, pauses or interruption management when compared to speech only.

One of the problems with VMC appears to be the way that timing can be disrupted; this may undermine basic communication processes (O'Conaill and Whittaker, 1997). In a comparative study of FTF, low quality VMC and high quality VMC, O'Conaill and Whittaker (1997) used speech process measures relating to speaker transitions (e.g. backchannels; handovers); if VMC/FTF differences were due to disruptions due to audio lags etc., this would be apparent in the low quality video condition but not in the high quality. However, results showed that the high quality video connection, predicted to be comparable to FTF, still led to more formal example of handovers etc. This suggests that it is not a simple technological issue (i.e. due to audio delay); rather, VMC was used in a different manner to FTF.

Whittaker (2002) notes that these findings show little reason to add a video channel to communication technologies; there may not be any added benefits. There may be exceptions to this; Ames et al. (2010) demonstrated that children prefer video chat to audio phonecalls, due to the possibility of showing the other speaker what is being held or looked at. Use of videophones will possibly cross over to remote collaborative work in a reversal of how technologies have been 'socially constructed'

in the past; for example telephones were thought to only be of use for business conversations and early computer applications were designed for simple office tasks (Postmes et al., 2000).

Many studies that investigate the effects of mediated communication do not clarify the behavioural processes that are being served by a given medium; for example, paralinguistic signals such as tone of voice can be conveyed over the telephone, and facial expressions can be seen over a video connection. It may be necessary to pay more attention to these fundamental communication behaviours in order to understand why a particular medium has demonstrated a particular result. The relevance of a behaviour is related to the task being undertaken; it may be that in problem solving tasks (e.g. Chapanis et al., 1972) there would be no advantage in transmitting facial expressions, but this may be different in tasks involving persuasion or negotiation. A further point is that all FTF interactions are not equal; Schegloff (1980) points out that there are many different types of person-person interaction, for example meetings, debates, interviews and ceremonies. These all have different norms for turn-taking and appropriate behaviours; this may have implications for the way that communication technologies affect this aspect of communication, and suggest that disruptions to turn-taking will only be an important factor depending on the task being undertaken.

"For some of the contemplated innovations, like computer conferencing, exchanges of letters may be a more appropriate past model to study, for there too more than one may 'speak' at a time, long lapses may intervene between messages, sequential ordering may be puzzling (as in 'Did the letters cross in the mail?') etc."
(Schegloff, 1980; p81.)

Schegloff argues that the possible danger of CMC is that the sequential ordering of talk may be disrupted; in this case it would be difficult to maintain intersubjectivity, as this is thought to occur through repair mechanisms relying on sequential organisation.

The examples outlined previously appear to show that there may be some advantage of process measures over outcome measures in measuring communicative efficacy. Outcome measures have not always been found to distinguish between conditions; although in some of these cases process measures have been found to vary (e.g. Anderson et al., 1997). This suggests that in order to understand communicative effectiveness examination of the process of communication may be more fruitful. The process – content and structure – of communication should illustrate the various functions that are needed despite the ungrammatical, non-sentential nature of dialogue. This approach of examining actual dialogue, rather than attempting to impose a grammatical/sentential framework upon it, solves the methodological problem assumed by Chapanis et al. (1977):

"The most important observation about these data is how exceedingly difficult it is to quantify natural human communications. Perhaps the greatest source of difficulty arises from their unstructured, unruly, and error-ridden nature. When they are examined closely, natural human communications, whether oral, written, or typewritten, appear at first glance to follow few syntactic or grammatical rules."
(Chapanis et al., 1977; p122.)

FTF is generally thought to be the most effective and efficient means of communication; does this seem to be the case from the evidence outlined? Or is it possible that in some cases other media are more appropriate; if so, under what conditions?

There is evidence that there is not always a difference between FTF and CMC; e.g. for 'cognitive'-type tasks (e.g. Chapanis et al., 1972; 1977). If this is so, it is important to note that using media technologies may not in itself affect an act of communication. This leads to the question of why differences are apparently found in much of the research; in order to answer this it may be useful to examine the relationship between the characteristics of a given medium and the processes/behaviours that constitute the communication. For example, a medium that supports synchronous feedback should facilitate shared understanding; a medium such as email should hinder this process. In a study carried out by Oviatt and Cohen

(1991) it was demonstrated that receiving instructions over a telephone led to a faster task completion time than a non-interactive audio recording of instructions. Incremental feedback in the interactive condition also avoided redundant descriptions and instructions that were found in the non-interactive condition. 'Social'-type tasks where a more interpersonal perception of attitude or affect are required may benefit from the addition of visual information.

Due to the primary function of speech, an audio medium such as the telephone should be of benefit in almost all tasks; adding other media characteristics generally does not seem to improve task performance. This may be why videophone type media do not seem to reliably improve task performance; it has been suggested that video is more useful in creating a shared environment (Whittaker, 2002).

2.2 Theories of communication media

The following sections will examine some theoretical approaches explaining differences between FTF communication and CMC. From an examination of the relevant literature, several theories can be seen to explain the communicative features listed above; the grounding (or collaborative) model of communication (e.g. Clark and Brennan, 1991; Clark and Schaefer, 1989; Clark and Wilkes-Gibbs, 1986), Media Richness Theory (MRT) (e.g. Daft and Lengel, 1984, 1986), Media Synchronicity Theory (MST) (Dennis and Valacich, 1993, 1999), Dialogue Games Analysis (DGA) (e.g. Carletta et al., 1996).

2.2.1 The grounding/collaborative model

The grounding (or collaborative) model of communication put forward by Clark and Brennan (1991; see also Clark and Schaefer, 1989; Clark and Wilkes-Gibbs, 1986; Clark, 1996) is based around an apparently simple supposition:

"All collective actions are built on common ground and its accumulation." (Clark and Brennan, 1991; p127)

The process behind this is based on shared understanding;

"In conversation, for example, the participants try to establish that what has been said has been understood. In our terminology, they try to ground what has been said - that is, make it part of their common ground. But how they do this changes a great deal from one situation to the next. Grounding takes one shape in face-to-face communication but another in personal letters. It takes one shape in casual gossip but another in calls to directory assistance."

(Clark and Brennan, 1991; p128)

Clark and Schaefer (1989) argue that common ground develops if "the contributor and his/her partners mutually believe that the partners have understood what the contributor meant to a criterion sufficient for current purposes." This is the 'grounding criterion' (although this belief may be erroneous; it is possible for a complete conversation to be played out with the participants at cross-purposes).

This model states that there are two phases in making a contribution in dialogue:

- (1) Presentation phase - A presents to B
- (2) Acceptance phase – B registers understanding

Understanding could be confirmed through the lack of 'negative' evidence (e.g. clarification requests) or positive evidence, which is preferred (Clark and Brennan, 1991⁶):

- (i) Demonstration (strongest)
- (ii) Acknowledgement
- (iii) Relevant next turn
- (iv) Continued attention (weakest)

Two main factors are argued to shape grounding; the purpose (or what the actors

⁶ This model is based upon dyadic dialogue; it does not explain how grounding may be reached in multi-party dialogue (Eshghi, 2009).

intend to accomplish) and the medium (the available characteristics of the given technology, and also their inherent costs). Another aspect of the grounding theory refers to the lack of perfect understanding in all communication; for example it may not be enough to simply ask a question, if that question has not been understood. Rather, there is a level of understanding that is reached between the actors, based on the current purpose of the communication. This level is referred to by Clark and Brennan as the 'grounding criterion'.

In the Clark and Brennan model there is an acceptance of the structures of turn taking and repair seen in the Conversation Analysis body of research (e.g. Sacks et al., 1974; Schegloff et al., 1977; Schegloff, 1982 etc.). However, contributions are seen as consisting of two phases; a presentation phase of the original contribution, and a secondary phase of the acceptance of that contribution. The implication is that there is a fundamental difference between attempting a contribution and actually achieving that.

How do people perceive that grounding has been achieved? One way is to look for negative evidence, for example seen in "I don't understand..." type utterances; but that according to Clark and Brennan is too inefficient in communication - rather, people look for positive evidence of grounding. These generally take one of three types; acknowledgements, relevant next turns or continued attention.

A further proposal in the grounding model is that of least collaborative effort (further outlined in e.g. Clark, 1996);

"The principle of least collaborative effort: In conversation, the participants try to minimize their collaborative effort - the work that both do from the initiation of each contribution to its mutual acceptance" (quoted in Clark and Brennan, 1991, p135).

As mentioned earlier, the grounding criterion depends on the current purpose of the communication; this grounding may take the form of references (e.g. pointing or describing) or verbatim content (e.g. repeating or spelling out). The criterion may not

be equal on both sides; for example an adult describing something to a child may have a different idea of when something has been sufficiently explained.

The above has given a brief outline of the grounding model of communication. It has potential to predict the effects of mediated communication. The concept introduced by Clark and Brennan to describe and predict differences is that of costs; the effort demanded by a medium to convey any given message. Clark and Brennan claim that this is where the strength of their model comes in; whereas previous studies have remarked upon differences found between media, the grounding model can account for these differences. Clark and Brennan argue that there are eight ‘constraints’ on grounding based on media characteristics; i.e. the characteristics that may or not be supported by a given medium. Constraints in this terminology are not meant in a negative sense, they are those aspects of a medium which are free to vary. For example, copresence is not a constraint of the telephone – as it is not supported it cannot vary.

Clark and Brennan’s constraints:

1. copresence (physical presence)
2. visibility (either talking heads or shared workspaces)
3. audibility (speech)
4. cotemporality (synchronous vs. asynchronous)
5. simultaneity (sending and receiving simultaneously)
6. sequentiality (enforced sequencing of turns)
7. reviewability (recording of messages)
8. revisability (private self repair)

Within any given medium, there are also costs relating to the media characteristics; below are those suggested by Clark and Brennan (1991), although they admit that this list is not comprehensive:

- | | |
|------------------|-------------------|
| 1. formulation | 7. asynchrony |
| 2. production | 8. speaker change |
| 3. reception | 9. display |
| 4. understanding | 10. fault |
| 5. start-up | 11. repair |
| 6. delay | |

Clark and Brennan claim that this model can explain previous results, such as in the choice of medium for a particular purpose (e.g. Cohen, 1984; Furnham, 1982; Finholt, et al., 1990). In all of these studies it is claimed that this model can explain media choice in terms of cost relative to purposes.

Convertino et al. (2008) state that the accumulation of common ground, built up incrementally through joint action, 'leads to a greater efficiency or a minimum effort for communication' (p. 1637). This raises the question of who is benefiting from this efficiency – does it also apply to overhearers or bystanders who may be part of a group collaboration? Convertino et al. draw a distinction between 'language as a collaborative activity' and 'cooperative work' carried out by workgroups. In order to utilise Clark's grounding theory in terms of workgroups, three differences between CMC and computer-supported cooperative work (CSCW) are suggested: firstly, cooperative work involves the formation /negotiation of social relationships as well as working procedures. Secondly, Clark et al. do not explain how common ground is built and utilised by groups of three or more rather than dyads. Thirdly, cooperation in groups is dependent on shared understanding of how the work will be carried out; this is the coordination of process rather than content. Convertino et al. present results from a study based upon Clark et al.'s theory which they claim demonstrate communication within workgroups becoming more efficient in terms of changes to the turn structure over time. The number of turns for the task increased and turns also became shorter (interruptions were also measured and remained stable). Further evidence was presented utilising 'Conversation Game Analysis' which identified relative frequencies of different dialogue acts such as 'check', 'clarify' and 'align' (see Kowtko et al., 1991). It was found that these turns that checked understanding increased over task runs for each condition. This is contrary to the idea that common ground increases understanding; these types of dialogue acts would be expected to decrease as common ground accumulates. Convertino et al. explain this in terms of i) the proportion of clarification requests increasing as the task efficiency increased due to their usefulness to the task and ii) information checking increasing as team members learn the importance of sharing task-relevant information. It is not clear if useful predictions can be made from this approach to Clark's theory due to the post-

hoc nature of the explanations given.

2.2.2 Media Richness Theory

Another theory that has been commonly referenced in CMC research is Media Richness Theory (MRT); 'richness' is defined as,

"The ability of information to change understanding within a time interval. Communication transactions that can overcome different frames of reference or clarify ambiguous issues in a timely manner are considered rich. Communications that require a long time to enable understanding or that cannot overcome different perspectives are lower in richness. In a sense, richness pertains to the learning capacity of a communication."

(Daft and Lengel, 1986; p. 560.)

Media richness is a function of four things; the medium's capacity for immediate feedback, the number of cues and channels available, language variety, and the degree of personal attention to the recipient. This leads to a gradient of communication technology richness going from FTF (richest) to simpler symbol systems such as numeric text or semaphore flag waving. Depending on the task, it should be possible to predict the level of 'richness' required for maximum efficacy, and select the medium that is most appropriate. This is related to the level of 'situational equivocality'; where multiple and conflicting viewpoints are required for a task, a higher level of richness is required. Uncertainty, or a lack of available information, would require a leaner medium in order to be most effective.

Media richness theory predicts an intuitively attractive relationship between task type and medium, illustrated in Table 2.

Table 2: Task-richness requirements according to Media Richness Theory
(adapted from Hollingshead and McGrath, 1993)

	Task type	Computer systems	Audio systems	Video systems	Face-to-face
Richness required increasing downwards	Generating Ideas and plans	Good fit	Marginal Too rich	Poor Too rich	Poor Too rich
	Choosing Correct Answer	Marginal Too constrained	Good fit	Good fit	Poor Too rich
	Choosing Preferred Answer	Poor Too constrained	Good fit	Good fit	Marginal Too rich
	Negotiating Conflicts of interests	Poor Too constrained	Poor Too constrained	Marginal Too constrained	Good fit
	Richness transmitted increasing from left to right				

Task types here can be seen as a continuum of increasing interdependence; negotiation in reaching consensus demands the greatest medium richness and therefore will only be suited to FTF communication. It may be that fitting a medium to a task is not this straightforward though; richness requirements may decrease over time in a stable group, as they adjust to the constraints of the medium and develop normative uses of the technology (McGrath and Hollingshead, 1993); for example the use of emoticons e.g. :-) for a smiling face to refer to an ironic or sarcastic comment in email. In view of this process it may be that over time adaptations to the given technology will have an effect of reducing the constraints of the medium; but it is not clear from MRT how this appropriation of norms would develop as a function of the group or the medium. MRT can be seen as a development of previous theories based around the support of social cues through media (e.g. Morley and Stephenson, 1969; Short et al., 1976). Whittaker (1995) argues that these theories are all based around the non-verbal communication hypothesis; if non-verbal signals such as head nods, facial expressions, eye contact or voice inflection are missing the flow of communication can be disrupted (Argyle et al. 1968; Kendon 1967; Rutter and Stevenson 1975). These theories, including MRT, are based upon the following assumption;

"Communication media affect group functioning in large part by the degree to which they transmit social context cues."
(Straus and McGrath, 1994; p88.)

MRT was originally developed in order to compare traditional media (e.g. telephone and written documents); it is not clear to what extent this model can account for differences in new communication technologies.

The Media Synchronicity Theory (MST) model emerged from the lack of experimental evidence for MRT. MST attempts to refocus attention on the underlying communication processes (Dennis and Valacich, 1999; Dennis et al., 2008). This is in contrast to the MRT approach of reducing the task to a single characteristic and then attempting to match that to an appropriate medium. This theory adapts McGrath's (1991) TIP (Time, Interaction, Performance) theory in terms of various processes (inception, problem solving, conflict resolution and execution) taking place within group functions; however, here McGrath's group well-being and member support functions are combined into a single 'social' function. The two group functions then, are social and production. Both of these functions may be affected by the medium used; for example communication relating to the task will differ depending on the familiarity of that task (McGrath, 1991). Communication relating to the social function will be affected by familiarity between group members; if they do not know each other the primary goals will be affected (e.g. McGrath and Hollingshead, 1993). If the group has previously been established it is likely that there will be group norms (Gersick and Hackman, 1990), and they will be able to carry out tasks with less rich information (McGrath and Hollingshead, 1993; Yoo and Alavi, 2001).

Dennis et al. argue that there are two processes that are universal for all groups regardless of task; conveyance and convergence. The conveyance of information comes from actions such as questions and proposals, triangulation of information (i.e. seeking confirmation) and contextualisation (i.e. relating the current group situation to past events). Following from a stage of deliberation, convergence of shared meanings is possible through affiliation, a stage of mutually agreeing upon meanings. Dennis et al. point out that the two central concepts in MRT, equivocality and uncertainty, are not relevant here; it is the process rather than a feature of the task that is examined. As mentioned earlier, these processes will vary with both the familiarity of the group members and the task; the dynamics of a newly formed group facing an unfamiliar task will be very different to a familiar group with a familiar task.

Although this model is claimed to centre on group processes, there is also a model of relevant medium characteristics, which Dennis and Valacich claim have emerged from the model of communication put forward by Shannon and Weaver (1949). This model draws distinctions between the source, transmitter, channel, receiver and destination; it is the channel that represents the medium and Dennis and Valacich claim that there are only five characteristics/'capabilities' that are important:

1. Immediacy of feedback
2. Symbol variety
3. Parallelism (simultaneous conversations possible)
4. Rehearsability (private self repairs possible)
5. Reprocessability (re-reading and questioning)

Dennis and Valacich note that these characteristics are not fixed; they may range from low to high. For example, in FTF communication symbol variety is usually seen as high; this may not be the case if there is not a common shared language or other impediment. Dennis and Valacich claim that this model allows three conclusions beyond those of MRT; firstly, that no one medium can be claimed to be the 'richest'; secondly, that medium capabilities are not fixed; and thirdly, that it is impossible to rank media in terms of their 'richness'.

Table 3: Characteristics of different modes of communication according to Media Synchronicity Theory (adapted from Dennis and Valacich, 1999)

medium	feedback	Symbol variety	parallelism	rehearsability	Reprocessability
Face-to-face	high	low-high	low	low	Low
Video	medium-high	low-high	low	low	Low
Telephone	medium	low	low	low	Low
email	low-medium	low-high	medium	high	High
Written mail	low	low-medium	high	high	High

A novel aspect of this model is its concept of synchronous communication; rather than solely being based on the immediacy of feedback, here this is tempered by the level of parallelism. A high level of feedback will not necessarily support individuals

working together if there are a large number of conversations occurring simultaneously. Dennis et al. (2008) suggest that a low level of synchronicity may be preferred for conveyance (exchanging information); a high level may be preferred for convergence (agreeing). Conveyance and convergence could also be affected by the other medium capabilities; reprocessability is thought to be advantageous for deliberations that may be needed in conveyance or convergence, as is rehearsability, but symbol variety is thought to be too closely tied to the task for any general claims to be made.

Dennis and Valacich propose that this model leads to predictions about the influence of media on group task performance, based on the level of synchronicity provided; high levels of synchronicity support convergence, and low levels support conveyance. For example, with an unfamiliar group, more convergence will be needed, as well as higher symbol variety; for familiar groups synchronicity is less important.

2.2.3 Dialogue coding approaches

The theories that are most amenable to quantitative testing are those which propose a coding scheme for annotating dialogue transcripts. Dialogue Games Analysis (DGA) (e.g. Carletta et al., 1996) came about as a coding scheme for analysing the HCRC Map Task corpus (the experimental paradigm used will be described in the next chapter). Carletta et al. (1997) claim that their approach has three advantages over previous dialogue coding schemes such as Walker and Whittaker (1990), Alexandersson et al. (1995) or Condon and Cech (1996). Firstly, they claim that DGA goes beyond previous attempts that have been dependent on the dialogue type being examined. Secondly, naming utterances according to their discourse role or goal only allows analysis at a basic level; DGA is claimed to be more useful in that it looks at dialogue on three different levels (see below). Thirdly, it is claimed that this approach can be used to examine any dialogue structure; it can also be used in conjunction with coding of other dialogue phenomena.

The DGA coding scheme breaks dialogue down into three levels; moves, games and transactions. Transactions are the dialogues which attempt to complete some goal; a part of a larger task. These consist of games which are made up of moves. Moves are the mutually exclusive components such as initiations and responses; games may overlap or be embedded. These are shown in Table 4 below.

Table 4: ‘Moves’ from Dialogue Games Analysis (Carletta et al., 1996)

Dialogue Game Moves		
1. Initiating moves	2. Response moves	3. other
Instruct	Acknowledge	Ready
Explain	Reply Y	
Check	Reply N	
Align	Reply W	
Query YN	Clarify	
Query W		

The DGA coding is achieved by applying a dialogue coding protocol to transcripts that have been organised according to the ‘turn’ distinction proposed by Chapanis et al. (1977), beginning when one speaker ‘takes the floor’ and ends when another begins speaking. A decision tree is applied to the turns; this firstly distinguishes between ‘initiation’, ‘response’ or ‘preparation’ and then leads to further questions to identify the correct dialogue move. For a description of how the DGA protocol is applied and examples of the various moves see e.g. Newlands et al., (1997); Newlands et al. (2003). A development of DGA is Dialogue Macrogame Theory; Mann (2002) acknowledges that this approach is most appropriate for task-oriented dialogue: “It is easy to find dialogues that DMT cannot represent. We expect that to account in a comparable way for natural dialogue as a whole, several other theories will be needed.” (Mann, 2002).

DGA is a way of labelling parts of dialogue purely in terms of function, or how the speaker is achieving their goals. It does not examine the linguistic properties and therefore can only be applied to task oriented dialogue where the goals of the speakers are transparent. As the design was created to account for the goals of the speakers undertaking the map task, when applied to further tasks it has to be considered whether additional (or even fewer) games and/or moves should be

introduced, for example: “The full set of Conversational Games described by Kowtko et al. (1991) was found to be necessary and sufficient to code the CMC interactions. No new categories of Games or Initiating Moves were required.” (Newlands et al., 2003; pp 336-337). That this is a necessary consideration for each separate application of the DGA protocol to a new set of transcripts demonstrates one limitation of this approach.

When applied to the contrasting media conditions in the HCRC map task, the difference between media was characterised as more ‘align’ and ‘check’ games in the audio-only condition; in the FTF condition “Participants in face-to-face context could make use of non-verbal signals (such as eye gaze) as well as verbal forms of grounding; hence face-to-face dialogues were shorter than the dialogues in the spoken context (Doherty-Sneddon et al., 1997).” (quoted in Newlands et al., 2003, p. 330). When applied to a face-to-face versus CMC comparison of the map task, it was found that different frequency profiles of the moves was found over the two conditions, with the difference thought to be due to convergence in the referring and instructing terms developed over CMC; it was suggested that these concise forms of communication are arrived at only through CMC and are not found in spoken dialogue (Newlands et al., 2003). Note that accuracy in the map task CMC condition improved over time, but only up to the level of accuracy consistently achieved by participants in the spoken condition. These studies demonstrate the usefulness of an approach that utilises coders applying a protocol to dialogue transcripts in order to extract various measures for quantitative analysis.

Although there are wider uses for a dialogue based coding system, DGA cannot be applied to all types of communicative acts. It seems to be most applicable to a dyadic, spoken, task-based dialogue. Naturally occurring dialogue cannot easily be broken down into identifiable goals as the interlocutors’ intentions or motivations may be completely opaque.

Another coding scheme closely related to DGA and also based upon the HCRC map task corpus is the Typology of Move Attributes (TMA) (Davies, 1998). This

approach attempts to identify both the strategies used by an interlocutor and those points where a particular strategy would have been expected but was not utilised. Davies (2006) tested four hypotheses based upon Grice's (1975) cooperative principle (plus a general folk-linguistic notion of cooperation), Clark's (1996) collaborative theory and Shadbolt's (1984) principle of parsimony. Applied to the HCRC map task data, Davies argues that evidence supports the principle of parsimony; individuals attempt to minimise effort at the expense of the task.

The DGA approach to dialogue annotation can be contrasted with another current approach, International Standard (ISO) 24617-2. This approach attempts to identify and label dialogue acts in a multidimensional and more formal manner than previous approaches (the coding protocols presented in this thesis require no prior knowledge of linguistics).

'Utterances in interactive discourse, such as spoken dialogue, have one or more communicative functions that characterize the type of communicative action which the participants are performing; these functions carry an essential part of the meaning of dialogue utterances. An adequate characterization of this aspect of meaning requires a coherent system of well-defined communicative functions. This standard provides empirically as well as theoretically well-motivated concepts for defining communicative functions, for identifying dimensions of interaction that dialogue acts may address, and for functional dialogue segmentation.' (ISO 2010 p. 2)

This approach attempts to create a classification scheme for the various functions of parts of dialogue that both human and computer coders can reliably apply to identify and label the functions that an utterance fulfils.

The International Standard for annotating dialogues is intended to draw together advances in dialogue annotation that have been developed independently and overlap in their categories; also the terminology used has been mutually inconsistent. This is due to the fact that they were intended for individual specific purposes, e.g. the DGA studies based upon the map task (Carletta et al., 1996); the TRAINS project (Allen et al., 1994); the Verbmobil project (Alexandersson et al., 1998). The currently proposed ISO standard builds upon these previous studies and others such as

DAMSL (Allen and Core, 1997) and DIT++ (Bunt, 2006; 2009).

The ISO standard identifies dialogue acts and provides a formal language for expressing these categories.

‘Dialogue acts are such actions as providing information, requesting the performance of a certain action, apologizing for a misunderstanding, and providing feedback... Distinctions such as that between a question and a reproach concern the communicative function of a dialogue act, which is one of its two main components. The other main component is its semantic content, which describes the objects, properties, relations, actions and events that the dialogue act is about. The communicative function of a dialogue act specifies how an addressee should update his/her information state with the information expressed in the semantic content, when (s)he understands the speaker’s utterance.’
(ISO 2010, pp. 5-6)

Identifying and labelling dialogue acts usually involves assigning a communicative function to some piece of dialogue (this is the approach taken in this thesis). Further detail can be provided by describing the semantic content, the proposed ISO standard distinguishes between several categories (‘dimensions’) such as feedback information, turn allocation information and topic progression information. Also included is information about previous dialogue acts which a subsequent dialogue act may be semantically dependent on, for example a question and answer.

It is claimed that existing dialogue annotation schemes take one of two approaches – either in terms of intended effects or in terms of the form of the dialogue act. Problems arise with the form when for example a question has the form of a statement (e.g. “You’re going home tomorrow”).

The proposed ISO standard consists of 9 core dimensions:

1. Task
2. Auto-feedback
3. Allo-feedback
4. Turn management
5. Time management
6. Discourse structure management
7. Social obligations management
8. Own communication management
9. Partner communication management

In a similar manner to the collaborative model, the proposed ISO standard attempts to identify the goals of the actors as well as the dialogue phenomena that supports the different interactional functions. This approach is most appropriate for computational linguistics but less useful for researchers into human-human interaction that do not share a background in linguistics. In contrast, the method used in this thesis identifies specific aspects of interaction that directly relate to mutual understanding and can be applied to any dialogue without any need for prior knowledge of the speakers' goals, intentions or information states.

Whittaker (1995) argues that communication can be seen to consist of two features: co-ordination of process, and content (e.g. Clark and Brennan, 1991; Clark and Schaefer, 1989). Conversational process refers to the management of the interaction, such as turn-taking so that only one person is speaking at once, and initiations and closings (e.g. Sacks et al., 1974). On the other hand, co-ordination of content refers to the maintenance of shared beliefs and understandings (Clark and Brennan, 1991; Clark and Schaefer, 1989) and interpersonal attitudes (Short et al., 1976). These two aspects are carried out through voice, gestures and other visual behaviour; it is thought that the restriction of these modes of communication lead to various effects in mediated communication. This dichotomy of process and content will provide the background to the coding protocols described in this thesis.

There is a separate observable dichotomy wherein participants in dialogue either implicitly or explicitly negotiate meaning in order to achieve 'intersubjectivity'; or

conversely participants feel that communication has been achieved successfully (or feel that the communication is inconsequential) and no questions or clarifications are needed. It is this latter case that gives rise to arguments over when or where people had agreed to meet for example; explicit negotiation of meaning is an important and necessary part of goal-oriented communication.

2.3 Quantification of mutual understanding in dialogue

Communication is possible through the construction of mutual understanding. This can only be achieved if references and the meaning of elided material within a contribution to dialogue have been established in a context that can be accessed by the parties involved.

One approach taken in this thesis is to find indices of ‘grounded’ material in dialogue. Healey et al. (2008) demonstrated the basic viability of quantifying ellipsis phenomena as a quantitative index of mutual-accessibility of context. Healey et al. demonstrated that the frequency of use of cross-speaker elliptical expressions in online chat varies systematically depending on whether communication is ‘local’ i.e. within a single chat room or ‘remote’. However, the coding of ellipsis in the (2008) study did not follow an explicit protocol. It relied mainly on the distinctions made by Fernandez et al. (2004) but specific measures of reliability and validity were not calculated. The ellipsis protocol presented here has been previously found to be reliable and can be applied to dialogue regardless of medium (Colman et al., 2008).

Dialogue creates and develops a shared context, e.g. through anaphoric reference. This process is known as ‘grounding’ (e.g. Clark, 1996) and allows the understanding of contributions that may be completely opaque to someone not able to access the context; i.e. they do not share the common ground.⁷ The common ground grows as it develops through processes such as repair (e.g. if the antecedent of an anaphoric reference is not successfully understood and needs explaining).

⁷ See e.g. Eshghi (2009) for evidence of side participants and overhearers attempting to access shared context through the analysis of ellipsis.

Examining the use of ellipsis in dialogue then, would appear to be one way of examining the extent to which participants are able to access and understand their shared context.

2.3.1 Accessing elided material in dialogue

“Speakers’ utterances normally contain discourse phenomena known as anaphoric references and ellipsis. Corpora of human-human conversations or of dialogues between a human and a simulated (Wizard of Oz) machine demonstrate the prevalence of these phenomena. Anaphoric references occur when a speaker refers back to something mentioned earlier in the conversation, e.g. ‘Read *that message*.’, ‘Are there any emails from *her*?’. Ellipsis occurs when a word or phrase is ‘left out’, but can be understood from what has gone before, e.g. ‘Are there any ... from Peter?’ where the noun ‘emails’ has been left out.” (Williams, 1996; p.1)

Spontaneously generated dialogue, whether naturally occurring or task-oriented, rarely sticks to accepted rules of grammar or even politeness. Interruptions, ungrammatical utterances and grunts or other noises are found in the majority of contributions in dialogue corpora. One reason for this is the ubiquitous use of ellipsis; the omission of words or phrases from a contribution which can be inferred or extracted from previous contributions. Ellipsis is optional; the full constituent could serve communication as well as the elliptical version. Where ellipsis occurs across speakers i.e., one participant makes (elliptical) use of another’s contribution, it provides an index of the mutual understanding of the conversational context (see e.g. Eshghi and Healey, 2009; Healey et al. 2008). In some cases this may seem obvious, as in the polar response ‘yeah’, signifying that a question has been heard, understood and considered; however, there are degrees of complexity which would seem to require a close understanding of what another participant is referring to. It is this particular aspect of mutual understanding that can be investigated through the quantification of elliptical phenomena.

The term ‘ellipsis’ as used by Colman et al. (2008); Healey et al. (2008); Eshghi and Healey (2009) is an umbrella term for *the omission of part of a contribution with the*

assumption that the other participants in the dialogue will be able to infer the full constituent meaning; “ellipsis occurs where some constituent is omitted from a turn and can only be resolved by reference to the content or syntax of a preceding turn”, Healey et al. (2008).

Healey et al. also used the term ‘ellipsis’ to refer to types of anaphora; a pronoun may replace a relatively complex noun phrase. The use of the term ‘ellipsis’ in this thesis should not be taken to mean that this work is a linguistic study of ellipsis in language or that all types of ellipses are identified in this approach. The approach taken in Eshghi and Healey (2009) and Healey et al. (2008) did not attempt to provide a taxonomy of the various types of ellipses or explain their usage in a linguistic sense; rather it was used as a term that encompasses those dialogue phenomena which can be seen to require access to a shared context of previously established words and phrases, or material that could be inferred. Elliptical contributions and anaphora are used to index the presumed accessibility of shared context. If people can use them (instead of the equivalent fully spelled out expression that they stand-in for) then it provides a measure of the extent to which they believe that they share that context:

“Ellipsis... indexes the extent to which the meaning of an utterance depends directly on the context of the preceding dyadic exchange i.e. the extent to which participants assume the common ground established during the dyadic exchange is accessible to each other.”
(Eshghi and Healey, 2009; p. 4)

Whenever dialogue is flagged as being problematic through the use of repair or a repair initiation (discussed below), it could be assumed that mutual understanding is not being achieved. Rather than accepting the absence of repair as positive evidence of mutual understanding, instances of shared context being accessed can be identified through the absence of words or phrases; understanding the contribution with elided material implies reconstruction of the full constituent through accessing the presumed shared context - grounded material that is *a positive index of mutual understanding*.

“Ellipsis provides a useful window on the accessibility of different contributions at different points in a dialogue... An elliptical expression involves the omission of one or more words or phrases from an utterance which are, by implication, presumed to be recoverable from the local context... Interpretability of elliptical utterances - as manifest in the relevant response to them - demonstrates that the context required to recover the missing bits (the antecedent) is sufficiently salient or pragmatically ‘close’ or ‘in focus’, for the responder.” (Eshghi and Healey, 2009).

The use of ellipsis in dialogue was touched on in a previous map task study, Newlands et al. (2003),

“Our study focuses on how these predicted changes in communicative strategy are achieved, examining adaptation to CMC both within individual turns of dialogue (e.g. in use of ellipsis, turn length, etc.) and across wider sections of dialogue (e.g. conversational games).” (Newlands et al, 2003; p. 328.)

The only result regarding ellipsis that was predicted was that ellipsis may be used more over time, making turns shorter; however visual inspection of transcripts showed this not to be the case. Also it was argued that ellipsis would have had a negative effect on performance:

“In addition, these reductions [in turn length] cannot explain the increase in performance, of course, because ellipsis and pronominal reference are harder to process than fuller descriptions.” (Newlands et al., 2003; p. 342.)

There is experimental evidence that demonstrates this is not always the case. Studies examining convergence in referring expressions show that descriptions of unfamiliar objects become shorter with successive uses (Krauss and Weinheimer, 1966; 1967). Recent experimental work (e.g. Mills, 2007; 2011) shows that pairs working together produce shorter descriptions over time and these increase success; if new pairs are formed (i.e. without the shared referring expressions) they perform worse. The descriptions that are converged upon can be both highly elliptical and use anaphoric references (Mills, 2011).

For the purposes of quantitative analysis, indices of presumed shared context have

been operationalised in this thesis as three main types: anaphora (exaphora, cataphora, endophora and vague anaphora); non-sentential utterances (NSUs) (broadly broken down into questions, answers, statements with subtypes of these three categories) and sentential ellipsis (sentences with elided material that is not a sentence fragment).

The following sections will describe the use of anaphora, non-sentential utterances and sentential ellipsis in dialogue, and present the coding protocol for indices of presumed shared context. The approach taken here is ultimately concerned with demonstrating the viability and usefulness of measurable indices of mutual-understanding of context; although this borrows from formal or computational linguistics it should not be thought of as belonging to those disciplines. Anaphora, NSUs and sentential ellipsis are introduced as a measure of mutual-understanding of context as in order for an ‘elliptical’ contribution to be comprehended there must be some shared knowledge, or ‘common ground’⁸.

2.3.2 Ellipsis: shared context in dialogue

Ellipsis has been defined as,

“an implicit reference to some material, either previously mentioned, or somehow inferable. In both cases, this elided material can be successfully recovered and applied to the remnant of the ellipsis, that is, to the piece of information stated explicitly in the elliptical fragment.”

(Alcantara and Bertomeu, 2005)

The precise usage of the term ‘ellipsis’ may be more difficult to universally define; various working definitions are used in different approaches. A general working definition used in a broader linguistic sense could be that there are three main types of ellipsis; sluicing (Ross; 1969), noun phrase ellipsis and verb phrase ellipsis. It is also possible to classify ellipsis as either ‘shallow’ or ‘deep’; shallow ellipsis can be resolved through an antecedent presented earlier in the dialogue⁹ whereas deep

⁸ If an elliptical contribution is not understood it would generally generate a repair sequence.

⁹ Cataphoric reference is an exception; the anaphoric reference is given initially.

ellipsis requires some prior knowledge that has not been explicitly mentioned (Frederking, 1993).

In the approach described and utilised in this thesis, ‘ellipsis’ is used to refer to three phenomena; anaphora, non-sentential utterances (NSUs) and sentential ellipsis and is referred to more generally as *presumed shared context*. It should be noted that anaphora are generally not considered to be elliptical in linguistic theory; here they are included as they require some level of shared knowledge in the same way that other forms of ellipsis do. These categories are divided further; the indices of presumed shared context coding protocol figures below show how these are defined and identified.

2.3.3 Shared context in dialogue: anaphora

In the current approach anaphora are categorized into four different types; endophor, cataphor, exaphor and vague anaphor.

Endophor is the most commonly recognizable of these phenomena; following a noun (e.g. an object or name; the antecedent) a pronoun is used to refer to it subsequently. For example, “Jack passed his exam!” “Oh, he did? Good for him!”.

Cataphor is similar, although in this case the pronoun occurs before the explicit reference, e.g. “Doctors thought he would die, but the window-cleaner survived the fall.”

Exaphor is generally a reference to some ‘thing’ in the immediate environment of the dialogue participants, e.g. “Pass that over here.”¹⁰

Vague anaphor is used as a category for anaphoric reference that does not refer to anything specific, e.g. “That’s it!”

¹⁰ This may not necessarily refer to an object in the physical environment.

2.3.4 Shared context in dialogue: non-sentential utterances

“the interpretation of non-sentential utterances or fragments, as they are traditionally called (e.g. Morgan, 1973)) is highly context dependent. ... the resolution of the intended content of fragments can be modelled as a by-product of the establishment of coherence in dialogue, which ... we define as the establishment of a meaningful connection of the content of the current utterance to its discourse context.”
(Schlangen and Lascarides; 2003).

Non-sentential utterances (NSUs) have been extensively studied previously (e.g. Schlangen, 2004; Fernandez, 2006). For the purposes of this approach a non-sentential utterance is defined as a dialogue contribution that does not contain a verb. The most common of these have previously been found to be polar responses, acknowledgements and one-word queries “Huh?”, “What?” (Fernandez, 2006).

There have been various taxonomies of NSUs; below in Table 5 is given that from Fernandez (2006) based upon an examination of the British National Corpus. Table 6 shows the NSU categories used in the current approach, outlined in detail in Chapter 3.

Table 5: Example taxonomy of Non-Sentential Utterances (NSUs) taken from Fernandez, 2006

Acknowledgements	Questions	Answers
Plain acknowledgement	Clarification ellipsis	Short answer
Repeated acknowledgement	Direct sluice	Plain affirmative answer
	Check question	Repeated affirmative answer
		Propositional modifier
		Plain rejection
		Helpful rejection
Extensions		Completions
Factual modifier	Filler	
Bare modifier phrase		
Conjunct		

Table 6: Non-Sentential Utterance taxonomy used in this thesis, taken from Colman, Eshghi and Healey (2008)

Answers	Questions	Statements
Polar answer	Sluice	Rejection
Acknowledgement	Clarification ellipsis	Modifier
Prompted NSU answer	NSU check	Continuation
Unprompted NSU answer	NSU query	

The NSU types used in the current approach (above) were decided upon following an examination of previous taxonomies, sample dialogues taken from the British National Corpus and a sample of HCRC Map Task transcripts. Rather than being totally comprehensive, it was intended to provide a tool for identifying the most common forms of NSU. It can be seen that some types have been utilised from previous approaches, e.g. sluice and clarification ellipsis. These two categories can both be thought of as non-sentential NTRIs; a simple sluice may be equivalent to a low specificity NTRI (e.g. “what?”) and clarification ellipsis refers to a non-

sentential repeat of previously contributed material used for clarification. This is not necessarily the case all of the time; there are examples of sluicing that may be equivalent to high-specificity NTRIs. The term ‘sluicing’ comes from Ross (1969); in the approach taken by Fernandez (2006) direct sluices are treated differently to clarification requests/NTRIs as they may be “not due to a communication trouble, but instead ask for further information that was explicitly or implicitly quantified away in the antecedent utterance.” (Fernandez, 2006; p.15). In this respect they are comparable with Schegloff et al.’s (1977) position two NTRI non-component referents. The two examples below from the BNC quoted by Fernandez (2006) illustrate this:

A: who did you interview?
 B: Benjamin
 A: **when?**
 B: last night.
 (BNC KEO 138-141)

A: Anyway Jim so you’re off to Australia?
 B: yeah
 A: **where?**
 B: er Melbourne.
 (BNC HVO 1015-1018)

2.3.5 Shared context in dialogue: sentential ellipsis

Sentential ellipsis differs from NSUs in that the utterance contains a verb, such as in verb phrase ellipsis e.g. “Jack can play the guitar and Mary can too.” In this case the non-elliptical version of this sentence would be “Jack can play the guitar and Mary can play the guitar too”.

2.3.6 Ellipsis and levels of participation

Goffman (1981) examined the relative status of direct addressees, side participants, overhearers and bystanders (DAs and SPs are ‘ratified participants’; Goffman, 1976). These levels of participation can be seen to demonstrate different levels of grounding, especially in mutual-understanding of context seen through elliptical reference resolution (e.g. Healey et al., 2008). Eshghi and Healey (2009) examine

ellipsis in multiparty interaction; specifically, what is the difference in understanding between people being directly addressed (DA), side participants (SP) and overhearers? DAs and SPs seem to have access to the same common ground in this study. Previous evidence came from Wilkes-Gibbs and Clark (1992); tangram studies show that side participants are more likely to use previously shared information than overhearers.

Healey et al. (2008) showed that ellipsis frequencies (using the Fernandez et al. (2004) taxonomy rather than the ellipsis protocol used in this thesis) demonstrate contrasting usage in ‘local’ and ‘remote’ virtual locations in a MUD. The questions of whether ‘overhearers’ have access to the common ground shared by direct addressees and side participants, and if participant structure alters significantly through the use of CMC was investigated using an analysis of ellipsis. It is clear that in some ways CMC differs; Healey et al. (2008) show that in the ‘Walford’ (text-based online talker) community participants engage in three concurrent conversations on average. In a comparison sample of FTF dialogue from the BNC the average was one. This leads to complex patterns of turntaking in Walford transcripts. ‘Remote’ virtual locations led to lower use of ellipsis than ‘local’ locations; however, this is not the case for dyads in which case there is no difference.

To recapitulate, although mutual-accessibility of context is fundamental to communication there has not been a reliable method for observing or measuring it. The ellipsis coding protocol presented in this thesis (Chapter 3; Section 3.3.2) provides a useful step in this direction. It gives a standardised coding scheme that can quantify the extent to which speakers can directly access the constituents of each other’s turns. The protocol used in this thesis defines ellipsis as anaphora (use of pronouns), non-sentential utterances (NSUs; contributions not containing a verb) and sentential ellipsis (e.g. verb phrase ellipsis). The NSU taxonomy was a development of previous research; previous work has found approximately 10% of turns in dialogue to be non-sentential (NSU) elliptical phenomena (Fernandez, 2006; Fernandez and Ginzburg, 2002; Thompson, 1980). The current ellipsis protocol simplifies this and includes further instances of shared understanding. Ellipsis coding

has previously been used to demonstrate levels of participation in dialogue; here it will index shared understanding. In order for contributions to be understood when they consist of sentence fragments or contain anaphoric reference there must be a sufficient level of mutual understanding – specifically mutual accessibility of shared context, also described as a part of ‘grounding’. This could be thought of as a pool of shared knowledge that must be accessible in order for dialogue to move forward without trouble. Colman, Eshghi and Healey (2008) presented a protocol for identifying instances of anaphora and ellipsis in dialogue which is utilised in this thesis to code the HCRC map task corpus.

2.4 Conversation Analysis

Conversation Analysis (CA) is a method of studying naturally occurring conversations in order to identify underlying common features; this is done through painstaking transcriptions of conversations in which very little is not notated in some way (see e.g. Atkinson and Heritage, 1984). Although many different features of conversation have been identified in this field, the most relevant for the examination of mutual understanding are those dealing with turn-taking and repair (e.g. Sacks et al., 1974; Schegloff et al., 1977; Schegloff, 1992; 1997; 2000). Although the processes underlying turn-taking may be relevant in certain media comparisons¹¹, this may not be the case in examples such as asynchronous media (e.g. email) or certain synchronous media (e.g. interactive graphical communication). The transcripts used for analysis were smoothed-out to remove interruptions and overlapping; analyses of this could not be included in this study. The most applicable findings for this study from the CA tradition are those dealing with the organisation and process of repair. Clark and colleagues (see e.g. Clark and Brennan, 1991) suggest that dialogue builds mutual understanding through the process of grounding; repair being one of the grounding mechanisms. The idea of ‘grounding’ being critical to interaction is used in e.g. computer dialogue systems research, however there is no consensus of what should be studied; for example Fernandez et al. (2007) examined

¹¹ Doerry (1995) found turn-taking to be a major trouble source leading to repair within a collaborative task using audio or audio-visual media contrasted with FTF.

‘push to talk’ type dialogue systems and measured as ‘repair’ clarification requests and ‘rejection or correction’, also ‘repair sequences’ that involved reattempting part of the task. Although a notion of ‘repair’ is thought to be important in dialogue research, there has not been a previous attempt to create a reliable measure of identifying repair in the CA sense. The previous CA work on repair has been qualitative or descriptive. In this thesis an approach is described which attempts to extract quantitative data from communicative acts in order to allow direct statistical comparisons.

2.4.1 Repair and repair initiations in dialogue

Schegloff (2000) offers a comprehensive explanation of what is meant by repair in the CA sense:

“By ‘repair’, we refer to practices for dealing with problems or troubles in speaking, hearing, and understanding the talk in conversation (and in other forms of talk-in-interaction, for that matter...”
(Schegloff, 2000; p. 207.)

“Finally, the sorts of actions underwritten by the practices of repair are not limited to ‘correction’, nor are their targets limited to ‘errors’ – hence the use of the terms ‘repair’ and ‘repairable’ or ‘trouble-source’. There can be ‘trouble’ grounded in other than mistakes... Whatever the response – whether modification/correction or confirmation/repetition/reaffirmation – the ongoing trajectory of the interaction has been stopped to deal with possible trouble, and that marks this interlude of talk-in-interaction as repair ...”
(Schegloff, 2000; p. 209)

The CA notion of repair can be distinguished from more specific types, such as correction, as it is not necessary for there to be any observable error for a repair to be carried out (Schegloff et al., 1977). In general, ‘repair’ refers to a ‘self-righting mechanism’ (Schegloff, 1992) that deals with problems in speaking, hearing or understanding talk; if perceived trouble pauses the forward momentum of the talk this trouble must be dealt with before the forward momentum of the talk can continue. A distinction is drawn between the initiation of a repair and its subsequent success or failure; repair may be initiated by either the speaker or others. Repairs or

repair initiations may be identified according to the position in which they occur; for example this may be the trouble source turn itself, a response to the trouble source, or a response to the response. These distinctions between ‘self’ and ‘other’, ‘repair’ or ‘repair initiation’ and position one, two or three are critical to the categories captured by the repair protocol used in the corpus analysis.

2.4.2 The validity of applying the CA model

The CA tradition is based around the qualitative analysis of naturally occurring conversations, and as such, it may seem invalid to apply CA principles to the statistical analysis of multi-modal communication. For example, Schegloff (1993) argues against the quantification of conversational phenomena, illustrating his position with an example of research into degrees of sociability; investigators contrasted ‘categories of participants with respect to “laughter per minute” and “backchannels per minute”’ (Schegloff, 1993; p. 104). Schegloff argues that there are various problematic areas with this approach; for example the claim that ‘per minute’ is a meaningful unit of conversation. Schegloff terms the units to be examined ‘environments of possible relevant occurrence’; ‘per minute’ is not a valid unit as within a given minute of interaction, it may not be relevant to laugh or exhibit other conversational phenomena. Another problematic area deals with the occurrences of the phenomena being counted within their possible relevant environment. Quantifying these occurrences involves establishing every possible form of occurrence; this may lead to invalid judgements being made due to the difficulty in defining or operationalising the phenomenon. For example, a contribution such as ‘Yeah’ could be taken to denote a backchannel, although it may have been intended to signal a change of speaker. Although these criticisms of quantifying communication are valid for many conversational phenomena, they may not invalidate the quantification of repair and repair initiations.

For example, in discussing other-initiated repair, Schegloff notes

‘In fact, because nothing can be excluded in principle from the class “repairable” (Schegloff et al., 1977, p. 363), such repair initiation by the recipient of some talk appears to be the only type of turn in

conversation with an unrestricted privilege of occurrence; it can in principle occur *after any turn at talk*. In that respect, then, its “environments of relevant possible occurrence” are well defined.’ (Schegloff, 1993; p. 115.)

He then remarks,

‘Furthermore, there appears to be a determinate set of turn formats used to initiate such repair (Schegloff et al., 1977, pp. 367-369), and those formats that are also used to accomplish other actions are specifiable, as are the other actions they are employed to enact. So what counts as an instance of other-initiated repair is relatively well-defined.’ (Schegloff, 1993; p. 115.)

Due to the nature of repair initiations, they can be performed at any time; this allows their relevant environment to cover any turn in conversation. Also, due to the considerable amount of work carried out by Schegloff and others in the CA tradition, there exists a detailed description of the features that characterise the various types of repair. This would suggest that for conversation at least, quantification of other-initiated repair phenomena does not involve the potential problems that can be identified with other conversational phenomena, as Schegloff admits:

‘Unlike the earlier discussed practices of reference to persons, it appears that this domain of practices of talking-in-interaction – other-initiation of repair and its sequelae – can be “qualified” for quantitative treatment.’ (Schegloff, 1993; p. 115.)

Schegloff concedes that both other-initiation of repair, other-repair and repair in position three are potentially appropriate for quantification.

This being the case, the same arguments appear to hold up for self-initiated repair in conversation (at a simplified level, self-initiated repair can be characterised as ‘I meant this’, whereas other initiation of repair can be characterised as ‘I don’t understand’).

Are self-repairs in position one suitable for quantification? And do they reflect some aspect of mutual understanding? Firstly, they can be identified when there is evidence that an edit has taken place; also they can take place during any

contribution to dialogue. For these reasons it would appear that all categories of repair and repair initiation in position one, two and three can be quantified for analysis. The relevance of position one self-repairs for mutual understanding is that they give an index of both the difficulty of producing a contribution and the effort being invested in making a contribution understandable for the recipient(s). This may be of importance in contrasting modes of communication.

This does not validate the claim that interactions through other modes of communication can be quantified in this way, for example gestures or graphical interaction. However, the ‘environments of possible relevant occurrence’ would appear to generalise, i.e. any communicative act can be clarified or questioned. The question as to whether occurrences of repair phenomena can be identified within non-conversational modes of communication is less clear; it is assumed that although repairs in these cases may take different forms, underlying structures such as self/other and position could be identified.

Heritage (2004; Heritage et al. 2007) argues that in the area of institutional talk there is an argument in favour of quantitative analysis of dialogue;

“Although the question of quantification has been controversial in CA... it is clear that a number of questions about the relationship between talk, its circumstances, and its outcomes cannot be answered without the statistical analysis of results... if particular features of institutional talk are to be connected to characteristics of the participants such as attitudes, beliefs and, perhaps most important, the outcomes of the talk, forms of measurement must be developed that permit the relevant connections to be made.”
(Heritage, 2004, pp 137-138)

Heritage gives examples of how turn design has been utilised in a quantitative manner, e.g. Clayman and Heritage (2002) examined White House press interviews and coded questions for various characteristics in order to examine historical trends. These could be seen to change over the last 50 years and through controlling for a time trend variable other factors were identified as predictors of adversarialness in questioning.

“The description of these historical trends and the contextual factors to which they are sensitive would not have been possible without quantitative analysis.” Heritage, 2004; p. 139

Heritage (2004) describes how other aspects of dialogue are suitable for quantitative analysis, e.g. negatively polarized questions should have a measurable effect of a ‘no’ response. This was explored further in Heritage et al. (2007) which examined an intervention in home visits by physicians. The use of a simple phrase (“Is there something else you want to address in the visit today?”) which is considered positively polarized (Heritage, 2004) had a statistically significant effect based upon surveys that followed up the visit. This was contrasted with “Is there anything else you want to address..?” which was statistically the same as the control condition (no intervention).

Conversation Analysis has identified ‘repair’ as being a ubiquitous feature of dialogue; here it is utilised as a negative index of mutual understanding. The CA description of repair has been utilised (Healey (1999); Healey and Thirlwell (2002); Healey, Colman and Thirlwell (2005)) in the development of a protocol for identifying repairs of various types¹². If any communicative act can be accurately transcribed, it should be possible to apply the protocol to each separate contribution and identify which, if any, type of repair or repair initiation has occurred. This protocol allows measurement of various indices of communicative efficacy for the media being examined; for example the difficulty of producing a contribution (position one repairs). The effort being invested to achieve ‘grounding’ in Clark and Brennan’s model (1991) may be reflected in the number of position two and three repairs/initiations; all things being equal, this total should be a reflection of how difficult it is to achieve mutual-understanding or intelligibility within the restrictions of that medium. Although the concept of communicative trouble has been suggested previously as a method of measuring communicative efficiency within and between media (e.g. Urquijo et al., 1993; Doerry, 1995), a fundamental problem has been the subjective nature of trouble identification;

¹² See also Themistocleous et al. (2009); Colman and Healey (2011) for applications of the coding scheme.

“The determination of whether a particular exchange does, in fact, constitute a breakdown in communication is an inherently subjective assessment based on the analyst’s effort to retrospectively reconstruct the communicative significance of the behaviours documented in the transcripts... it is impossible to define deterministic, context-independent heuristics for recognizing breakdown based either on abstract features of the interaction (e.g. timing of utterances) or on the specific content of interaction (i.e. specific phrases or words).”
(Doerry, 1995; p88)

Although this may be true of Doerry’s approach, these criticisms do not hold up against the approach described here. There is reliability evidence to show that repair types can be identified by separate coders; this is due to position and type being identified through highly explicit instructions. It is not ‘impossible to define’ if e.g. a clarification request is a repair initiation (NTRI) in position two; it should also be clear that these identifications of repair types are ‘context-independent’. However, the approach described here agrees with the assertion that repair and communicative trouble phenomena cannot be identified through the presence of specific words or timing of conversational events; all judgements as to whether a contribution comprises a repair or repair initiation must be made by the individual coder.

2.4.3 Next turn repair initiations (NTRI)

The term ‘Next Turn Repair Initiation’ may be misleading; although Schegloff et al (1977) identified the ‘next turn’ as the most common position for a repair initiation, this may be delayed. For example, an abrupt change of subject may cause the repair initiation to be delayed while a new, ‘nested’ topic is dealt with and concluded before the previous topic of conversation is renewed (Schegloff refers to these and other less frequent types as ‘Delayed NTRIs’). For this reason the term ‘NTRI’ is used even though in some cases it will not appear to be the directly following on ‘next turn’.

“It is in fact the case that the vast majority of other-initiated repairs (some 90 per cent in the subsample I examined for this report) are initiated in the turn following the one in which occurs the trouble-source which the repair initiation is targeting.”
(Schegloff, 2000; p.211.)

Although Schegloff recommends that the term ‘NTRI’ not be used due to the implied location of the repair initiation, here it is used for continuity with Healey (1997);

Healey, Colman and Thirlwell (2005); Colman and Healey (2011). For this reason note that 'NTRI' refers to other-initiation of repair in general, almost all of which occurs in the next turn.

Schegloff, Jefferson and Sacks (1977) describe a hierarchy of NTRI types, based around how specific the repair initiation is. There are five levels suggested:

level 1	a non-specific signal that a contribution has not been heard or understood: 'huh?' 'what?' (lowest specificity)
level 2	question word: 'how?'
level 3	partial repeat plus question word: 'all the what?'
level 4	partial repeat
level 5	candidate understanding: 'you mean..?' (highest specificity)

The category of repair initiations referred to as 'non-component referents' (see directly below) by Schegloff et al. (1977) generally fall into the level 2 or 3 category. This can be seen because they do not necessarily reflect a high level of understanding of the contribution being acted upon (the trouble source), although a partial repeat may be included.

2.4.4 A special case of NTRI: Non-Component Referents

In some instances the structure of trouble source – initiation - repair may be ambiguous as to whether a repair initiation was intended. Typically these cases involve a question word and could be perceived as either the initiator assuming that they have missed some antecedent information, or conversely the initiator may be asking for further information and there has been no trouble. As the surface structure is the same in both cases, they both fall under the repair category 'non-component referents'. This form of NTRI is described by Schegloff et al. (1977) in the following way:

"There is a separate class of other initiators - in large measure using an

overlapping set of lexical items - which locate as repairables referents which were not actually components of prior turn:

Ben: They gotta - a garage sale.

Ellen: Where.

Ben: On Third Avenoo.

Ava: I wanted t'know if ya got a uhm whatchamacallit uhm p(hh)ark(hh)ing place this morning.

Bee: A parking place.

Ava: Mm hm.

Bee: Where.

Ava: Oh hh just anyplace heh heh I was just kidding ya."

(Schegloff, Jefferson and Sacks, 1977; p.369)

2.4.5 Disfluencies and self-repairs

Nicholson (2007) examined map task studies (not the HCRC map task corpus) in a study of disfluency (immediate self-repairs) in task-oriented dialogue, addressing whether the demands of the dialogue affected the amount of disfluency. Nicholson compared two contrasting theories of disfluency; a 'cognitive burden' view and a 'strategic modelling' view. The former sees disfluencies as accidents caused by the demands of the task; the latter as signals which strategically convey a difficulty to the listener. Nicholson concluded that there was evidence for both theories; deletions may occur for planning reasons and repetitions occur for hesitation reasons. This broadly corresponds with the position one repair distinction made in the repair coding protocol (articulation-formulation dichotomy, see Section 3.4); repetitions being articulation repairs (supporting the cognitive burden view) and deletions being formulation repairs (strategic modelling).

Experimental studies by Brennan and Schober (2001) examined whether disfluencies ('uh' and immediate self-repairs) affect comprehension. Instructions to select a particular coloured shape on a display were given to listeners; three different types of disfluency were manipulated and their reaction time measured. The three disfluent conditions were mid-word interruptions ('yel- purple square'), mid-word

interruptions with fillers ('yel-, uh, purple square') and between-word interruptions ('yellow- purple square'). Mid-word interruption with filler was found to lead to the quickest response. Not completing the word to be replaced is less misleading and leads to fewer errors, and the filler gives time for the listener to cancel the erroneous information. Although this may explain reaction time, it is unclear whether these results inform repair studies. The comparable repair category to examine is the position one formulation repairs. If there is a comprehension benefit in mid-word interruptions with fillers, this may show itself in the frequency of those repairs in the map task data. However, reaction times measured in milliseconds are a very different measure than those applied to the map task. Position one formulation repairs did exhibit a relationship with accuracy, but more follower P1F repairs indicated a less accurate route. Bortfeld et al. (2001) examined task oriented dialogue across factors of age; task difficulty; task role (director vs matcher); familiarity; gender. Disfluency was found to be highest with difficult task and 'director' role, "confirming that disfluencies are associated with an increase in planning difficulty" (2001, p.123). There were differences in the distribution of disfluency types; fillers were thought to be playing a more important role in interaction. The relationship between repair and disfluency in the current coding scheme is discussed further in Section 3.4.6.

2.4.6 Continuations and third turn repair

One category of NSU subtypes (Section 2.3.4 above; outlined further in Section 3.3) is 'continuation'; an utterance that follows on from something already contributed. This has obvious similarities with repair in position 3, and especially 'third turn repair' (Schegloff, 1997). The difference between the two comes in the specific definitions: an elliptical continuation must be non-sentential, i.e. not contain a verb; a position 3 repair must alter or amend a previous contribution. However, in usage it appears that the two are used in similar instances – intuitively it can be seen that a pause for a continuer can lead to a re-evaluation of the previous utterance. For example, the data presented in Chapters 4 and 5 allows the relationship can be seen from the highly significant correlation; Spearman's $\rho = 0.331$, $P < 0.001$ (follower and giver continuations correlated with P3 SI SR repairs). Further relationships from

tests of correlation and multiple regression analysis are examined in Chapter 6.

2.5 Quantifying mutual understanding

A quantitative method of identifying the components of communication that index mutual-understanding in dialogue is possible. This has previously only been attempted in a qualitative, descriptive manner through Conversation Analysis. Participants in dialogue (especially goal-oriented dialogue) often explicitly negotiate in order to achieve transparent mutual-understanding of the salient topic. Problems in understanding may be resolved through signalling a problem or attempted correction or modification, known as 'repair'. Unproblematic dialogue could be assumed when interlocutors successfully use elliptical or anaphoric expressions; missing explicit words or phrases are inferred by the intended audience. This approach will be discussed further in the next chapter.

In the grounding model, it was claimed that communicative outcomes are the result of costs affecting constraints on grounding mechanisms. This may be difficult to simplify into general predictions (due to the context sensitive nature of costs), but it would seem that media characteristics that allow participants to reach the grounding criterion through least effort would be more efficient, if purpose is held constant. It may be that those media that can more easily support feedback, or demonstrate continued attention and relevant next turns would be expected to be more efficient.

A problem with Clark and Brennan's model seems to be the difficulty in measuring levels of costs and benefits; it may be that it only allows post hoc explanations of results. Whittaker et al. (1993) point out that even in cases where costs can be operationalised, there may be a confounding factor of effort. In this case, a less efficient medium (entailing higher costs) may lead participants to invest more effort, concealing differences in communicative process if only outcome measures are examined. A problem with identifying this number of constraining characteristics is that it becomes hard to discriminate between the relative weight of factors. For example, would it be better to have audio-visual channels or the ability to review and revise your messages? The advantages of certain characteristics would seem to be

dictated by the task (or purpose); however, if their relative importance cannot be quantified it would be difficult to use these to formulate predictions.

Media Richness Theory suggests that media differences are due to the level of social cues being transmitted; this is due to the processes that are supported through non-verbal communication such as conveying attitudes of emotions. Task types are characterised in terms of equivocality or uncertainty; for the former, rich media will be most effective, lean media for the latter. Although this theory has been influential, it may be too simplistic to explain differences in new communication technologies; this is reflected in the lack of experimental evidence.

Media Synchronicity Theory looks at group processes, and brings in influences of familiarity of task, group and medium. This theory allows two central predictions; when there is a need for convergence, a high level of feedback and a low level of parallelism would be most efficient. For conveyance, a low level of feedback and a high level of parallelism would be preferred; in both of these cases high levels of rehearsability and reprocessability will increase efficiency. As with the Grounding model, it appears that the concepts involved may be difficult to operationalise. Although concepts such as task and group familiarity may be useful in contrasting effects, the relative importance of the various factors in this model are not made clear, making predictions difficult. For example, does rehearsability or revisability outweigh the level of synchronicity? A large part of this theory is based on parallelism, but in their analysis of media the only one they claim to have high parallelism is written mail (it is not clear why they think that parallelism would be lower for email or text chat). The implication is that FTF, video and telephone are very similar in the processes that they support, but this may not be the case. Another drawback to this model is the lack of any supporting experimental work; although Dennis et al. (2008) claim that the results of a previous study (Markus, 1994) are better explained through their model, post-hoc explanations should not be considered valid support in the absence of confirming data.

2.6 Review

Several theories have been outlined that explain the relationship between interaction and medium. Grounding/collaborative theory explains that common ground is built up through constant monitoring of shared understanding. This is constrained by the medium. Media Richness Theory argues that a medium may convey too much or not enough information; depending on the task the medium will help or hinder. Media Synchronicity Theory suggests that familiarity with the medium and other people are important; this affects the ability to put forward questions and proposals that lead to mutual agreement. Dialogue Games Analysis is a way of coding interactions in terms of the intentions of the interlocutors. This is examined at three different levels. This approach is more suitable for task-oriented dialogue than naturally-occurring dialogue. It was designed for analysis of the HCRC map task corpus, as was the Typology of Move Attributes.

This thesis describes a quantitative method of identifying the components of communication that index mutual-understanding in dialogue. This has previously only been attempted in a qualitative, descriptive manner through Conversation Analysis. Participants in dialogue (especially goal-oriented dialogue) often explicitly negotiate in order to achieve transparent mutual-understanding of the salient topic. Problems in understanding may be resolved through signalling a problem or attempted correction or modification, known as 'repair'. Unproblematic dialogue could be assumed when interlocutors successfully use elliptical or anaphoric expressions; missing explicit words or phrases are inferred by the intended audience. This thesis approaches mutual-understanding through the creation of reliable tools for extracting indices of mutual-understanding of context based on repair and ellipsis/anaphora.

Chapter 3

Methodology

3.0 Overview

This chapter outlines the approach taken in this thesis and describes a method for extracting quantitative data from dialogue transcripts that allows for direct comparisons in terms of the negotiation of mutual-understanding. Two separate approaches are taken; identifying positive indices of mutual understanding through the identification of presumed shared context (types of anaphora and ellipsis) and negative indices of mutual understanding through the identification of different forms of repair and repair initiations. This study is corpus-based and uses coding protocols to identify presumed shared context and repair phenomena; the frequency distributions across experimental conditions in the corpora can then be compared and contrasted.

3.1 General methods

3.1.1 Methodology: experimental corpora

Two sources will be used for illustrating the distribution of both repair and ellipsis; the British National Corpus (Burnard, 2000) and the HCRC Map Task Corpus (Brown et al., 1984; Anderson et al., 1996).

3.1.2 The British National Corpus

The British National Corpus (BNC) (Burnard, 2000) is a collection of examples of spoken dialogue and written British English that was collected from 1991–1994. The spoken part of the corpus, used in this study, was collected by volunteers and covers a wide range of personal and social characteristics. Spoken BNC transcripts are unscripted informal conversations from a demographically balanced sample. Full

details can be found at the BNC website.¹³ BNC sample extracts were used to establish baseline scores for both repair and ellipsis. For example, the BNC sample utilized to give a baseline measure of repair phenomena consisted of 30 dialogues made up from 1943 utterances/14725 words, provided by 41 people.

3.1.3 The HCRC map task

The map task was developed in order to elicit very clear examples of information transfer between two people. The original experimental paradigm was described by Brown et al., (1984). Detailed descriptions of the HCRC map task experimental design and outcome have been previously published, e.g. Boyle et al. (1994); Anderson et al. (1996). In pairs of participants one is randomly assigned to the role of ‘information giver’, and must describe a route to an ‘information follower’. Both participants have a ‘map’ with simple landmarks; but the two maps, although compatible, are slightly different in the presence or absence of landmarks. The ‘follower’ must listen to a route described by the ‘giver’ taken from a line drawn on the giver’s map and draw that route on their own follower’s map. Examples of the ‘giver’ and reciprocal ‘follower’ maps are given in Figure 1 below. The HCRC corpus of map task dialogues is used for this study as the dialogues illustrate a collaborative task in which each role is equally important; the giver cannot ‘transmit’ the route to the follower as in the Shannon-Weaver (1949) model; there must be negotiation of meaning in which both parties have equal responsibility.

There are two factors that were manipulated within the HCRC Map Task; eye contact and familiarity. In the ‘eye-contact’ or face-to face (FTF) condition the two participants were able to see each other over the angled boards that held their maps; in the ‘no eye-contact’ condition a screen was fixed between them that prohibited any possible eye contact or mutual gaze. ‘Familiarity’ was manipulated by recruiting subjects in familiar pairs; studies were carried out in ‘quads’ of two pairs who did not know each other. Studies could then be run with both familiar pairs and pairs of

¹³‘The British National Corpus’. <http://www.natcorp.ox.ac.uk/corpus/index.xml>. Retrieved 21st May 2012.

unfamiliar partners. A further factor that could be examined was that of task ‘role’; the differences between the dialogue contributions made by the information ‘giver’ and ‘follower’.

The HCRC map task corpus consists of 128 scripts; 19,133 turns and 156,315 words. Boyle et al. (1994) describe some preliminary findings of media differences from the corpus; using parametric ANOVA analysis Boyle et al. conclude that FTF is more efficient – due to fewer turns and words used in the FTF condition; also in the FTF condition there were more words per turn. Familiarity differences were found in that familiar pairs produced more turns and more words; there were no interaction effects between the two factors of medium and familiarity. Boyle et al. defined ‘backchannels’ as turns consisting of ‘uhuh’ or ‘mmhmm’ type acknowledging noises; more backchannels were found in the audio-only condition.

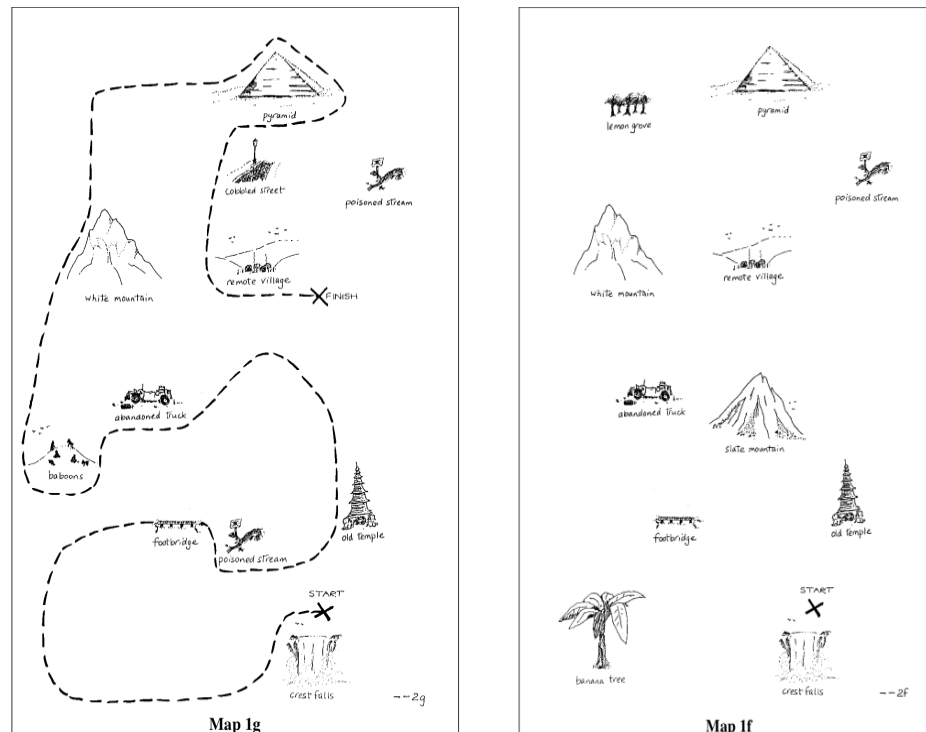


Figure 1: giver and follower maps

The giver map contains the route which must be verbally described to the other participant; it is not visible to the other participant. The follower map would have the route described drawn upon it.

The transcripts of the map task dialogues utilised here were collated in such a way that giver and follower turns are equal; interruptions and speaking over the other were 'tidied up' into a practical (for analysis) order. No examination of turn-taking phenomena has been undertaken for this thesis. The dialogues were between two people but multi-party dialogue with a complex array of direct addressees and bystanders would also be suitable for the mutual understanding analysis presented here.

3.2 Statistical methods

3.2.1 Methodology: IV and DV

For the map task transcripts three independent variables are used in these analyses: medium (eye contact or audio-only); familiarity (familiar or unfamiliar); task role (map route giver or map route follower). Independent variables such as sex or age which were collected as part of the HCRC map task data were not used. Dependent measures used for the analysis were: word counts; time taken; number of turns; task accuracy (measured as deviation from the given route); frequency measurements of the following: seven categories of repair/repair initiation, five levels of next-turn repair initiation (NTRI) specificity, four categories of anaphora, eleven categories of NSU/sentence fragments and one category of sentential ellipsis. The categories of dialogue phenomena to be quantified are shown in the table below. For the BNC baseline these dialogue phenomena were counted along with number of turns and words.

Table 7: Dependent variables: categories of dialogue phenomena to be quantified and used in the analysis

Repair and repair initiations		
1. P1 Articulation	2. P1 Formulation	3. P1 Transition Space
4. P2 OIOR	5. P2 NTRI	6. P3 OISR
7. P3 SISR		
NTRI specificity levels		
1. NTRI level 1 LOW	2. NTRI level 2	3. NTRI level 3
4. NTRI level 4	5. NTRI level 5 HIGH	
Anaphora categories		
1. Endophor	2. Cataphor	3. Exaphor
4. Vague Anaphor		
NSU categories		
1. Polar Answer	2. Acknowledgement	3. Prompted NSU Answer
4. Unprompted NSU Answer	5. Sluice	6. Clarification Ellipsis
7. Check	8. NSU Query	9. Rejection
10. Modifier	11. Continuation	
Other ellipsis categories		
1. Sentential Ellipsis		

A glossary of these terms with their abbreviation and a short description can be found in Section 3.3.1 below (also see Appendix A for a complete list of indices of mutual understanding).

3.2.2 Methodology: reliability of coding schemes

This section will illustrate the process whereby a coding scheme is decided to be sufficiently reliable for use. 'Reliability' in this case is used to refer to the degree to which two coders applying a coding scheme independently agree over their total decisions. As with many statistical measures, there is some debate as to the correct way to examine these data. The approach taken here is that suggested by Carletta (1996) and uses the non-parametric Cohen's (1960) kappa statistic to demonstrate agreement. This statistic is important because it factors out the chance agreement that

would have occurred, making this a stronger measure than percentage agreement. Carletta (1996) points out that many studies of dialogue use various measurements of reliability, which interferes with comparisons between studies. Carletta argues that it would be more appropriate to adopt Cohen's Kappa statistic. However, this issue has been re-examined recently (e.g. Carletta, 2008) and further approaches may be more accurate; in its favour Cohen's kappa is thought to be highly conservative i.e. less chance of a type I error ('false positive'). A further reason for using Cohen's kappa here is that this is the statistic most commonly associated with inter-rater reliability and therefore allows more close comparisons with the reliability of related coding schemes.

One problem with the Kappa statistic is the need for mutually exclusive categories in coding; naturally-occurring spoken dialogue especially does not lend itself to easily delineated units. As with DGA, the approach taken here is that mutually exclusive units can be identified. The simplest identifiable unit is a 'speech turn'; any contribution can contain many different types of repair or ellipsis phenomena which can be identified individually as separate and mutually exclusive entities¹⁴. It should be noted that the term 'speech turn' may be misleading; the coding protocols developed for measuring mutual-understanding have been designed without any particular communicative mode (e.g. speech, gesture, writing) in mind and are applicable to multiple modalities.

Cohen's Kappa statistic provides a measurement of agreement that takes into account expected random agreement; that which would be found if the dialogue was coded randomly. Kappa is calculated through the following formula:

$$K = \frac{P(A) - P(E)}{1 - P(E)}$$

where P(A) is the proportion of times the coders agree (agreements/agreements

¹⁴ That is not to say that a contribution could not contain many nested instances of both repair and elliptical phenomena.

+disagreements), and $P(E)$ is the proportion of times that the coders would be expected to agree if they were applying the codes randomly. There are further Kappa statistics developed for more than two coders (e.g. Fleiss' Kappa); Kappa statistics given in the following chapters are Cohen's Kappa calculated using the procedure given in Howell (1994).

3.2.3 Methodology: inferential statistical tests

Much work on dialogue and discourse takes a qualitative approach; in contrast, this thesis formulates metrics that will allow for inferential statistical analysis. Nonparametric statistical tests are less common in experimental work; they are highly conservative and therefore less likely to show a significant result i.e. they may be susceptible to type II errors but less prone to type I errors. This can be seen as a strength; a significant result using a nonparametric test is more likely to be evidence for a robust phenomenon. The approach taken here is to utilise both nonparametric and parametric tests (with appropriate data types – ratio or interval level).

3.2.4 Methodology: nonparametric statistical procedures

Nonparametric methods are usually used when there is some doubt that standard parametric techniques would be valid. For example, Likert scales (e.g. agreement on a scale of 1-5) can be ranked, but the data cannot be treated as interval or ratio level.

The major disadvantage of nonparametric techniques is that it allows fewer conclusions to be drawn from the data; e.g. the sign test may show a difference between two groups but the extent of this difference and confidence level is not shown. This is due to the fact that the data is reduced from its actual values to ranks; the order of the data.

3.2.5 Methodology: statistical correlations

In this thesis results from correlational tests will be presented; it is necessary to understand the implications of spurious correlations, confidence intervals and type I errors.

Spurious correlations are correlations that are due mostly to the influences of "other" variables. For example, as children get older their verbal and numerical skills increase; they also get taller. Despite the positive statistical relationship, it is not height that causes an increase in skills. Statistical correlations may suggest how relationships between variables work, but inferring cause and effect can be hazardous.

The second point to be aware of here is the confidence level; generally experimental papers assume that a significant result is produced through an inferential test providing a result with a 5% or lower chance of a type I error. This is not always the case; e.g. Doerry (1995) claims that when using a nonparametric test a 10% confidence level is appropriate. The confidence level is crucial when multiple tests are being carried out; many tests may allow many errors.

A further point to be acknowledged is that there is a qualitative difference between rank correlation coefficients such as Spearman's, and Pearson's product-moment correlation coefficient. The former is not mathematically comparable to the latter; a rank correlation coefficient is a measure of association rather than a linear relationship. This is not to say that nonparametric correlations are invalid in any way; they are used in this thesis when parametric tests are inappropriate.

3.2.6 Methodology: raw and adjusted data

There will be cases where two levels of analysis are possible; firstly, the raw totals of frequency counts and secondly, those totals adjusted for some covariate. For example, familiar pairs of participants may have longer utterances and more words than unfamiliar ones. Due to this, they may demonstrate higher frequencies of some verbal phenomena; this can be adjusted for by calculating a ratio of (phenomenon):

words¹⁵. A ratio score of this type also allows for the use of parametric tests (subject to appropriate conditions).

3.3 Positive indices of mutual understanding: presumed shared context

3.3.1 Glossary of terms

The coding protocol shown below was originally published in Colman, Eshghi and Healey (2008) where it was referred to as the ‘ellipsis protocol’. For the purposes of this thesis the term ‘presumed shared context’ is synonymous with the term ‘ellipsis’ as used in the Colman et al. paper and is to be contrasted with the negative indices of mutual understanding shown through repair and repair initiations.

Tables 8.1 and 8.2 list the individual categories that make up the indices of presumed shared context; the abbreviated term and a short definition is given.

The indices of presumed shared context consist of five main categories, with various subcategories. The first category is anaphora; the second non-sentential answers; the third non-sentential questions; the fourth non-sentential statements; and lastly sentential ellipsis.

The term *presumed* shared context is used due to the use of contributions utilising pronominal reference or elided material rather than giving the full constituent term. Such a contribution is thought to be made only when the ‘missing’ information can be inferred from the common ground, or shared context. If this material is not part of the common ground; the contributor was incorrect in presuming that that part of the context was shared; this state of affairs would be identified through the emergence of a repair sequence. For any ‘elliptical’ contribution, the contributor must presume that the shared context is sufficient to allow the intended full constituent to be understood.

¹⁵ E.g. (giver P1ART) divided by (giver WORDS); this approach also used by Louwse et al., 2007.

Table 8.1: Glossary of positive indices of mutual understanding: indices of presumed shared context

Type	Abbreviation	Description
Endophor	END	A pronoun with an identifiable antecedent
Cataphor	CAT	A pronoun used before the referent is specified
Exaphor	EX	A pronoun without an antecedent referring to something in the immediate environment
Vague Anaphor	VA	A pronoun used without an antecedent that does not refer to something in the immediate environment
Polar Answer	POL	Yes or no answers
Acknowledgement	ACK	A signal of understanding
Prompted NSU Answer	PA	A non-sentential answer that was requested
Unprompted NSU Answer	UNPA	A non-sentential answer that was not solicited
Sluice	SLU	A non-sentential question containing a wh- word
Clarification Ellipsis	CE	A sentence fragment that contains a repeat of a previous contribution due to trouble in understanding
Check	CHK	A query if an other participant has understood a contribution
NSU Query	NSUQ	A sentence fragment asking for more information

Table 8.2: Glossary of positive indices of mutual understanding (continued)

Type	Abbreviation	Description
Rejection	REJ	A sentence fragment that disputes a previous contribution
Modifier	MOD	A sentence fragment that describes or qualifies a previous contribution
Continuation	CONT	A sentence fragment that continues directly from a previous contribution
Sentential Ellipsis	SENT	A complete sentence that could be extended by including a previously used phrase

3.3.2 The ‘ellipsis’ coding protocol

The coding protocol shown below was originally published in Colman et al. (2008). A series of binary yes/no questions allow the identification of different types of anaphora, non-sentential utterances and sentential ellipsis. The categories captured by this protocol were judged to be the most commonly encountered through a series of different iterations of the protocol applied to excerpts from dialogue transcripts.

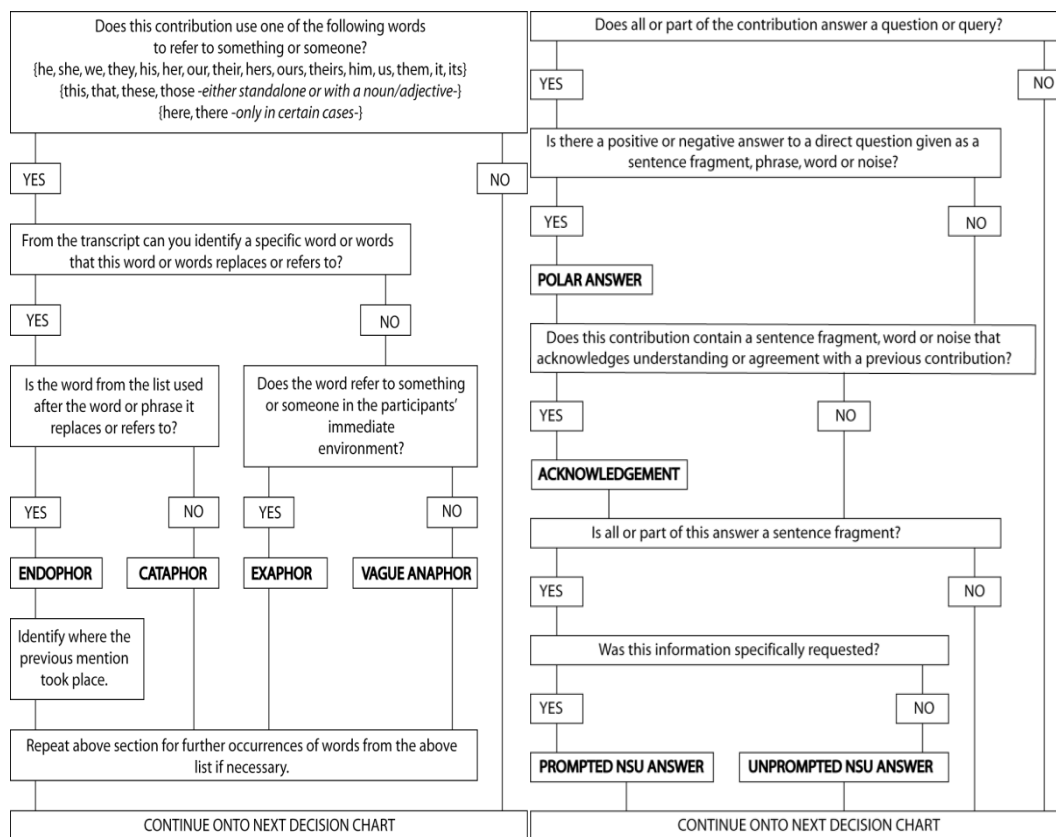


Figure 2: Indices of presumed shared context/‘ellipsis’ coding protocol (anaphora and answers)

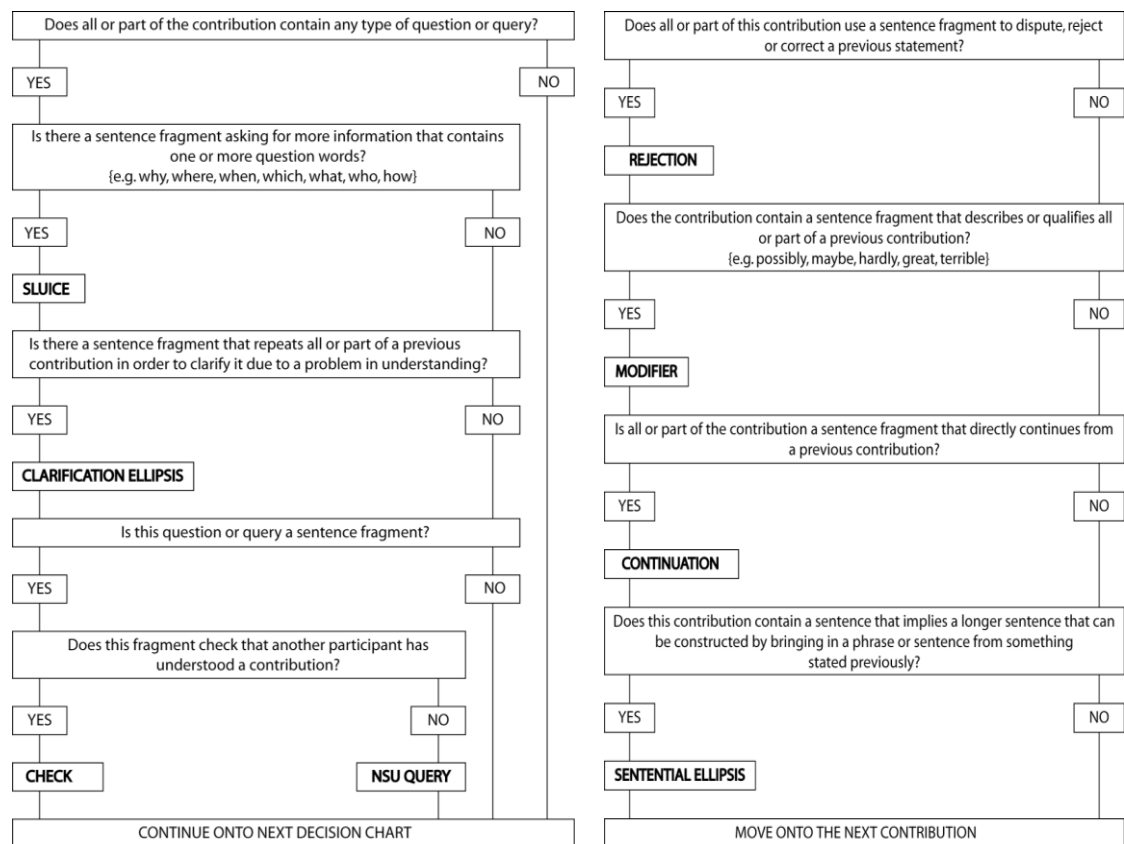


Figure 3: Indices of presumed shared context/‘ellipsis’ coding protocol (questions and statements)

3.3.3 Application of the presumed shared context coding scheme

The ‘presumed shared context’ coding scheme was applied to the entire HCRC map task corpus; results from this study are given in the next chapter. The protocol is used by taking a transcript of dialogue and examining each contribution individually. The protocol was designed to allow quick identification of anaphora, NSUs or sentential ellipsis; more detailed questions then identify the specific category. The coder asks themselves each question from the next box in the protocol with regard to the current contribution; in the task-oriented dialogue of the HCRC map task the author found that there are almost no instances of complex nested phenomena within a single contribution. In the coding of the entire HCRC map task corpus there were found examples of every category given in the protocol although some were more prevalent

than others.

Exemplars of the individual categories from the coding of the entire HCRC map task corpus can be found in Appendix B.

3.3.4 Reliability of the ‘indices of presumed shared context’ coding scheme

In order to demonstrate reliability, two coders (one computer scientist, one psychologist from the Queen Mary Computer Science department) applied the ellipsis protocol to a sample from the map task. The longest of these dialogues was chosen to be coded (transcript Q1NC1) which consisted of 446 turns and 5533 words. Kappa in this instance was .81, which shows high reliability. A confusion matrix of these results (see below) shows that there is a sufficiently strong agreement between the coders in most categories.

Table 9: Indices of ‘presumed shared context’ coding protocol confusion matrix:

	END	CAT	EX	VA	POL	PA	UNPA	ACK	SLU	CE	CHK	NSUQ	REJ	MOD	CONT	SENT	NONE	TOTAL
END	119		5	4													3	131
CAT	1	2																3
EX	1		8															9
VA	8	3	8	33													1	53
POL					113			5									2	120
PA						1												1
UNPA							7	1										8
ACK					5			78					2				8	93
SLU									2									2
CE										7								7
CHK											20	1					1	22
NSUQ							1			1		27			1		2	32
REJ													2				1	3
MOD														1				1
CONT							2					1			13			16
SENT																13	10	23
NONE			4	2		1	2	3	2	1			2		2	3	78	100
TOTAL	129	5	25	39	118	2	12	87	4	9	20	29	6	1	16	16	106	624
F(e)	27.08	.02	.36	3.31	22.69	.003	.15	12.97	.01	.1	.71	1.49	.03	.002	.41	.59	16.99	
Sigma f(e) = 86.92 sigma f(o) = 524 kappa = 0.81																		

For the anaphora categories there is strong agreement in the two most common categories endophor and vague anaphor. The other two anaphora categories had low

agreement but also low occurrence which would make them easier to miss or confuse with other anaphora types. For the NSU category types a similar pattern emerges; the most frequently occurring categories have high agreement whereas the rarer categories (e.g. NSU rejection) are disputed (2 instances agreed upon, 1 instance claimed by only one coder and 4 instances identified by the other, of which 2 were not identified and 2 were claimed to be an NSU acknowledgement by the other. Visual inspection of the coded transcripts showed that this apparently odd disagreement was due to ‘yes, we have no bananas’ type responses). The category ‘sentential ellipsis’ was highly disputed, with one coder identifying 10 instances that were not coded by the other (13 instances were agreed upon). Inspection of the transcripts showed that this was due to one coder not spotting instances of sentential ellipsis rather than confusing the instances for another category; for the example transcript used here the relatively few instances of sentential ellipsis were missed amongst the more frequent NSU types. This category was not removed as it was thought that awareness of the hard-to-spot nature of this category would ensure subsequent identification.

3.4 Negative indices of mutual understanding: repair and repair initiations

3.4.1 Glossary of terms

Table 10 below gives a breakdown of the categories used in the protocol identifying instances of repair and repair initiations. The coding protocol presented in Figure 4 differs slightly to that presented in Healey et al. (2005); repair types that were of very low frequency in preliminary coding of the HCRC map task (such as P2 Self-Initiated Other Repair and P2 NTRI Incomplete) were removed. Seven repair types were identified as being sufficiently common to be useful for statistical comparisons.

Table 10 gives the category with its abbreviation and a short definition in each of the seven cases.

Table 10: Categories of repair and repair initiation identified by the coding protocol

Repair type	Abbreviation	Description
Position one, self-initiated Articulation repair	P1A	An immediate self-edit of a contribution that does not change the meaning e.g. repetition.
Position one, self-initiated Formulation repair	P1F	An immediate self-edit that changes the meaning e.g. word replacement.
Position one, self-initiated Transition Space repair	P1TS	A self- edit or amendment to a contribution that takes place after a possible speaker change.
Position two, other – initiation of repair	P2NTRI	A signal that there is trouble in understanding an other’s contribution.
Position two, other – initiated other - repair	P2OIOR	A proposed revision or correction of an other’s contribution.
Position three, other – initiated self - repair	P3OISR	A self-revision or amendment to a speaker’s earlier contribution following an other-initiation.
Position three, self – initiated self - repair	P3SISR	A self-revision or amendment to a speaker’s earlier contribution without an other-initiation.

3.4.2 Construction of the repair protocol

The repair protocol given here for identifying the most common troubles in mutual understanding underwent several iterations before the current definitions were finalised. The original version of this coding scheme was published in Healey (1999). Further iterations were then presented in Healey and Thirlwell (2002) and Healey et al. (2005). For example, the following was the original wording of the position one decision box:

Does the initiator alter or amend their contribution during production?

Following application of this definition to exemplars taken from CA papers e.g. Schegloff, Jefferson and Sacks (1977) a change was made:

Does the initiator edit, amend, or reprise part of their contribution before another participant responds to it?

Further exemplars from the CA literature were examined until the current wording was decided to sufficiently capture this type of repair. The same process was carried out for all repair and repair initiation definitions in positions one, two and three.

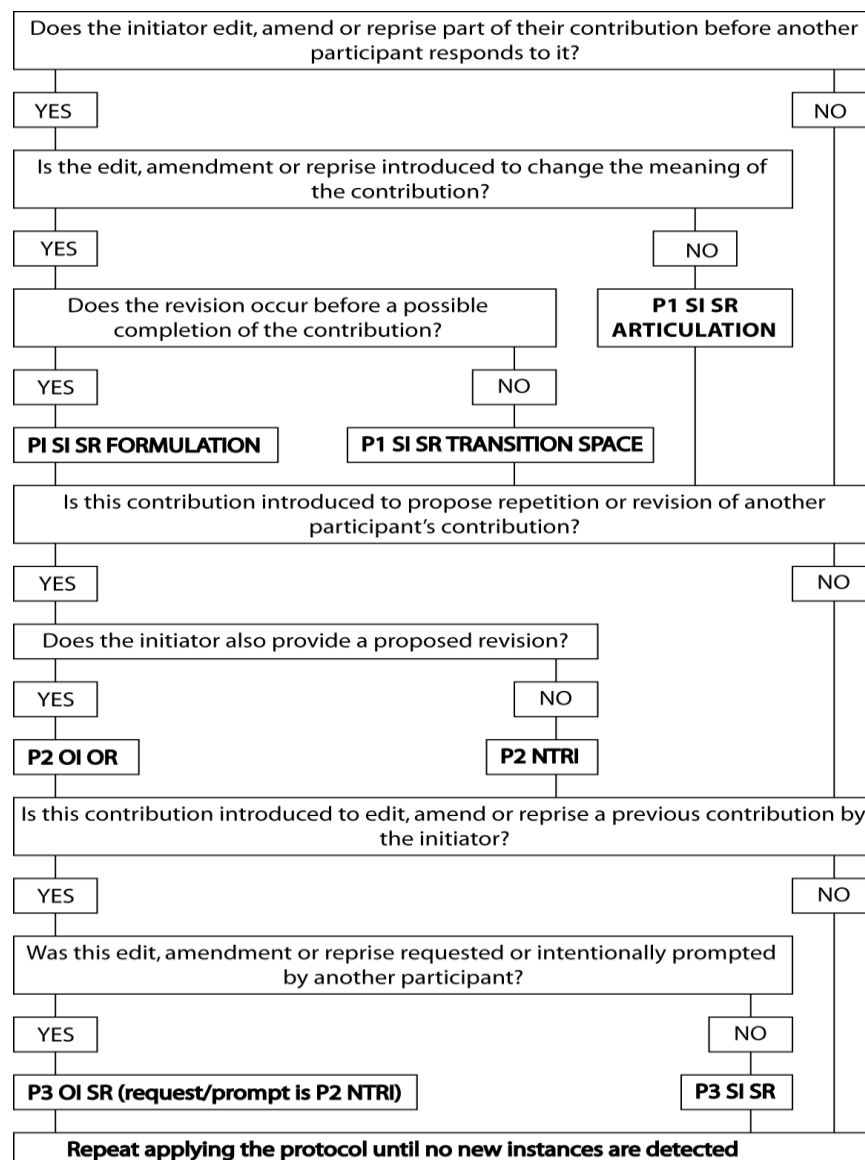


Figure 4: The ‘repair and repair initiation’ coding protocol

3.4.3 Application of the repair protocol

The protocol is used in the same way as the ‘ellipsis’ protocol presented previously. A series of yes/no questions are applied for each individual contribution to the interaction. The coder applies the protocol recursively to each contribution occurring in an interaction until no further repair phenomena can be identified.

Appendix C gives exemplars of each category in the coding protocol, captured through applying the coding scheme to transcripts of the entire HCRC map task corpus.

3.4.4 Repair protocol reliability

In order to examine the reliability of the repair coding protocol, two naïve users (not involved with the protocol design) were recruited from the Queen Mary Computer Science department and asked to apply the coding scheme to an unpublished corpus of doctor-patient consultancy transcripts. Reliability was calculated using Cohen's kappa. The confusion matrix from the two coders is given below. Kappa was calculated as 0.73 over 1728 comparisons which was judged to show sufficiently high reliability.¹⁶

Table 11: Indices of 'repair and repair initiation' confusion matrix

	P1A	P1F	P1TS	P2NTRI	P2OR	P3SISR	P3OISR	none	total
P1A	283	6	1					15	305
P1F	28	274	6					69	377
P1TS	2	4	4					12	22
P2NTRI				5				2	7
P2OR				2	6			7	15
P3SISR						34	1	13	48
P3OISR					1	1	2	1	5
none	13	54	9	4	3	30	8	828	949
total	326	338	20	11	10	65	11	947	1728
F(e)	57.54	73.74	0.26	0.05	0.09	1.8	0.03	520.08	
$\Sigma F(e) = 653.59 \Sigma F(o) = 1436 \text{ kappa} = 0.73$									

The above table shows the confusion matrix of two coders' agreements and disagreements for the different categories of repair which were used to calculate Cohen's kappa. Firstly, it can be seen that position one repairs are not identified in position two or three. Likewise, position two and three are not disputed; it can be assumed that the protocol identifies position reliably. Within a given position there is

¹⁶ Experimental validity for the protocol was examined in Healey et al.; 2005.

some dispute over the repair type. A further issue is where one coder has identified a repair and the other has not. For position one there is relatively little disagreement for articulation repairs; formulation and transition space repairs are more problematic. For formulation repairs one coder identified 69 instances that were not identified as repair by the other coder; 54 instances were identified by that coder which were not identified as repair by the other. However, the overall number of position one formulation repairs (377 versus 338 for the two coders) with 274 direct agreements shows sufficient reliability.

There appears to be difficulty in identifying when a transition space takes place. One coder identified 22 T.S. repairs, of which 12 were not recognised as repair by the other coder. That coder identified 20 T.S. repairs, of which 9 were not recognised by the other. Position one transition space repairs were not reliably coded, presumably due to their low frequency compared to other position one repairs. To a lesser extent the same problem occurs with P2NTRI, P2OR, P3OISR and P3SISR. There is a high level of agreement of where there is no repair suggesting that it can sufficiently be identified where a repair occurs, but there are disagreements over the specific type. Overall the kappa of 0.73 is considered to show sufficient reliability (Howell, 1994).

3.4.5 Uncoded CA repair phenomena

The repair protocol is deliberately limited in terms of the repair phenomena that it captures, loaded towards the most common. For example, there is no examination of the success or failure of a repair (Schegloff et al., 1977). Fourth position repairs are not coded; their occurrence is of such low frequency that statistical comparisons would not be useful. There is a type of repair which is of high frequency but also not identified as a separate type; third-turn repair (Schegloff, 1997). This type of repair is captured by the protocol as Position 3 self-initiated self repair (P3SISR), although the repair type has more in common with Position 1 self repair. It generally takes the form of a continuation, alteration or modification of speech that has been broken into (sometimes, but not always in a transition space) by another participant offering a continuer or other comment on what has been said.

For example (taken from Schegloff, 1997; p. 34):

Louise: I read a very interesting story today.
 Mom: Uhm what's that.
 Louise: W'll not today, maybe yesterday, aw who knows when hu- it's called Dragon Stew.

3.4.6 Repair or disfluency?

Disfluencies (e.g. Bard and Lickley, 1997) appear to have a surface resemblance to self-repairs, but the relationship between the two has not previously been investigated. Whereas the CA approach that has influenced the current methodology structures repair according to position/self vs other/initiation vs repair, there have been other attempts to classify 'repair' as an utterance containing a disfluency. Nicholson (2007) claims that Levelt's (1983) classification of 'original utterance' – 'editing phase' – 'repair' is important because it allows identification of different regions of disfluent speech, although she notes that other disfluency structures have been put forth eg Nakatani and Hirschberg (1994); Shriberg (1994). It is clear that disfluencies are ubiquitous, whether or not they share the same function as position one repairs;

“During spontaneous speech, speakers face the difficult task of rapidly deciding what to say and how to say it. One consequence of the speed at which speakers are required to do this is that they are often disfluent, producing phrases that repeat or correct what has already been said, as well as meaningless interjections such as um and er.” (Collard, 2009, p.1)

Bard et al. (2001) investigated whether 'disfluencies' (referred to as 'simple disfluencies'; repetitions; insertions; substitutions; deletions) are due to difficulty in planning, production, comprehending previous turns or other factors. Using as dependent measures disfluencies in the HCRC map task¹⁷ it was examined whether they behaved like the inter-move interval (IMI), the time difference between speakers. A multiple regression of predictor variables showed that disfluency does

¹⁷ Disfluency codings for the HCRC maptask can be found at <http://groups.inf.ed.ac.uk/maptask/deviation-scores.txt>. Retrieved 22nd May 2012.

not act like IMI, with, “length in words alone accounting for over 30% of the variance in total disfluency rate, more than all other groups combined” (Bard et al., 2001; p99).

By our definition, self-repair in position 1 and disfluency are not identical although there is some crossover. I can show that the phenomena that are commonly grouped together as ‘disfluencies’ are two discrete phenomena. Although both articulation and formulation repairs positively correlate with ‘disfluencies’ as coded for the HCRC DGA coding the two position 1 repair types have been shown to be separate. In the confusion matrix given previously for two naive coders, P1A and P1F are clearly distinguished from each other with 557 agreements and only 40 disputes over the P1 category. It will be shown in Chapter 6 that follower P1 formulation (P1F) is a predictor of accuracy; P1A is not.

Table 12: Position one ‘repair’ categories correlated with ‘disfluency’ codings from the HCRC map task

P1 repair	Pearson r	p
Correlation with ‘giver disfluency’		
Giver Articulation	.862	<.001
Giver Formulation	.858	<.001
Giver TS	-.138	.278
Correlation with ‘follower disfluency’		
Follower Articulation	.948	<.001
Follower Formulation	.877	<.001
Follower TS	.132	.298

From Table 12 above it can be seen that transition space repairs are a separate phenomena; they are not correlated with disfluency. If there are two separately identifiable types of repair phenomena being labelled with the term ‘disfluency’ it is not surprising that studies such as Nicholson (2007) could not identify their function.

The repair protocol presented here has been found to be reliable when used by naive users. The protocol identifies repair types in terms of self/other repair or repair initiation, and position; current contribution (position one), a following contribution

(position two) and a response¹⁸ to a position two contribution (position three). In order to provide a baseline measure of the relative frequency of different repair types, this thesis included the results of applying the protocol to 30 transcripts of naturally occurring dialogue from the British National Corpus. These results demonstrate in a quantitative manner claims originally made by Schegloff, Jefferson and Sacks (1977); specifically, that there is a preference for self-repair. When another participant has a problem with hearing or understanding there is a tendency to signal that there is a problem through the use of a repair initiation (often in the turn directly following the trouble source, 'NTRI' (Schegloff et al., 1977)) and allow the original contributor to self-repair. The frequencies captured through the protocol should allow an interaction to be analysed in terms of the sources of trouble and the strategies that are used to fix these. The most important distinction may be whether a trouble source is identified by the originator or other.

3.5 Review

This chapter has presented a methodology for extracting reliable quantitative measures of mutual-understanding in dialogue. There are two ubiquitous features of dialogue; the use of 'ellipsis' and 'repair'. Some types of immediate self-repair are also classified as disfluencies. Examining repair illustrates where trouble has emerged and how it is dealt with by the dialogue participants. The identification of elliptical contributions (including anaphora) gives a further measure of mutual-understanding. The correct inference of omitted words demonstrates that understanding has been achieved. If elliptical expressions are not understood then a repair sequence will ensue.

Reliability of the 'presumed shared context'/'ellipsis' and 'repair and repair initiation' coding schemes was shown in confusion matrices of two coders' agreements and disagreements. The two coding schemes can be seen to have strengths and weaknesses with respect to reliability within certain coding categories. Overall the inter-rater reliability was sufficiently high although some categories were

¹⁸ In some cases a re-evaluation or change of mind by the speaker rather than a response.

not reliably coded in this study, for example ‘sentential ellipsis’. The most commonly occurring categories had high inter-rater reliability.

The two confusion matrices demonstrating inter-rater agreement show that the disagreements are proportional rather than raw frequencies; the sheer number of position one self-repairs meant that there were many instances that were agreed upon. It is likely that with practice a higher reliability would be reached – but one strength of the protocols is that they are intended to be usable by researchers with little prior knowledge of linguistics; practice is not intended to be necessary.

Predictions from this approach are based upon the relationship between the medium and a communicative process. Whereas position one repair may signify a difficulty in expressing a contribution within a medium, position two NTRIs and position three OISRs would suggest a higher level of communicative coordination. The use of anaphora would also suggest that there are not problems with coordination or mutual understanding, although if repair initiations are high this may negate any positive outcomes.

Chapter 4

Positive indices of mutual understanding in the HCRC map task corpus

4.0 Overview

This chapter deals with a proposed measure of mutual-understanding in dialogue; that of presumed shared context or ‘ellipsis’ (Colman et al., 2008). For illustrative comparison a sample of the BNC is also coded. It is shown that anaphoric reference occurs more often in naturally-occurring dialogue. Non-sentential elliptical phenomena are highly prevalent in the map task transcripts and can in many cases be seen to occur more often than in naturally occurring dialogue. Different distributions of ellipsis are found in the different experimental conditions.

4.1 Method

This chapter presents both a comparison of the BNC sample and the HCRC map task coding and a breakdown and analysis of the HCRC map task corpus data. Firstly, the baseline measure of indices of presumed shared context in a BNC sample is contrasted with the overall profile of those indices within all 128 transcripts from the HCRC map task corpus. Secondly the occurrences of those indices in the map task corpus are presented in detail. Differences across the factors of medium, familiarity and role are given. SPSS software was used for analysis.

4.2 Results

Firstly the map task distribution is contrasted with the BNC sample. Medium, familiarity and role differences are tested (ANOVA) using frequencies adjusted for words.

4.2.1 Descriptive statistics

Of the 19,133 turns in the HCRC map task corpus, 4617 turns were found not to contain elliptical phenomena; 76% of turns were coded as containing elliptical

phenomena from the ellipsis protocol.

Figures 5 and 6 below illustrate the distribution of elliptical phenomena in a sample of the BNC and the entire HCRC map task. Visual inspection shows that the distributions are different, most clearly in the lower use of anaphora in the task oriented dialogue. Table 13 below gives the exact figures of the mean frequency (adjusted for word totals) of all indices of presumed shared context in both the BNC sample and the entire HCRC map task corpus.

Table 13: Exact figures for BNC and map task 'indices of presumed shared context' coding

	end	cat	ex	va	pa	ack	pnsua	unpnsua
BNC	.0379	.0111	.0496	.0202	.0059	.0131	.0020	.0013
Map task	.0218	.0051	.0004	.0096	.0217	.0332	.0034	.0012

	slu	ce	chk	nsuq	rej	mod	cont	sent
BNC	.0026	.0000	.0000	.0039	.0020	.0007	.0026	.0007
Map task	.0016	.0014	.0034	.0065	.0005	.0007	.0045	.0030

4.2.2 Overall distribution of indices of presumed shared context

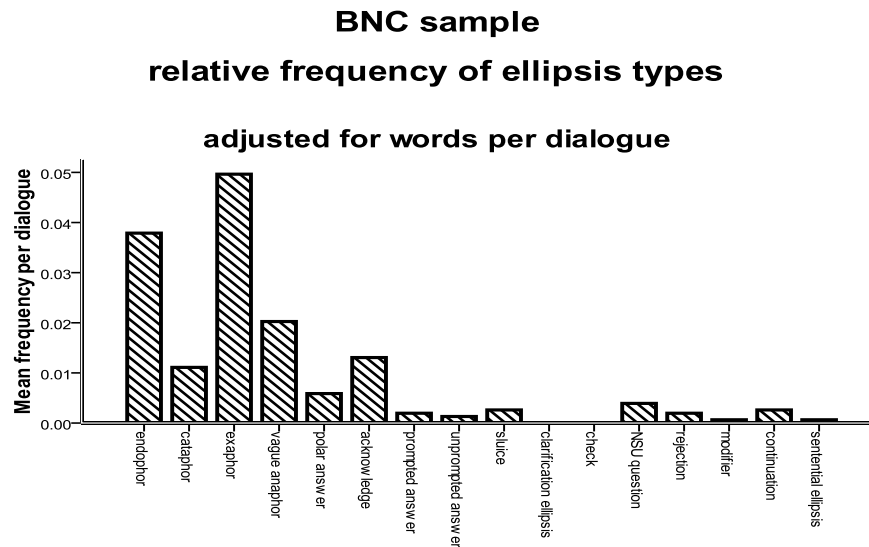


Figure 5: Frequency of ‘presumed shared context’/‘ellipsis’ categories in a sample of dialogue from the BNC

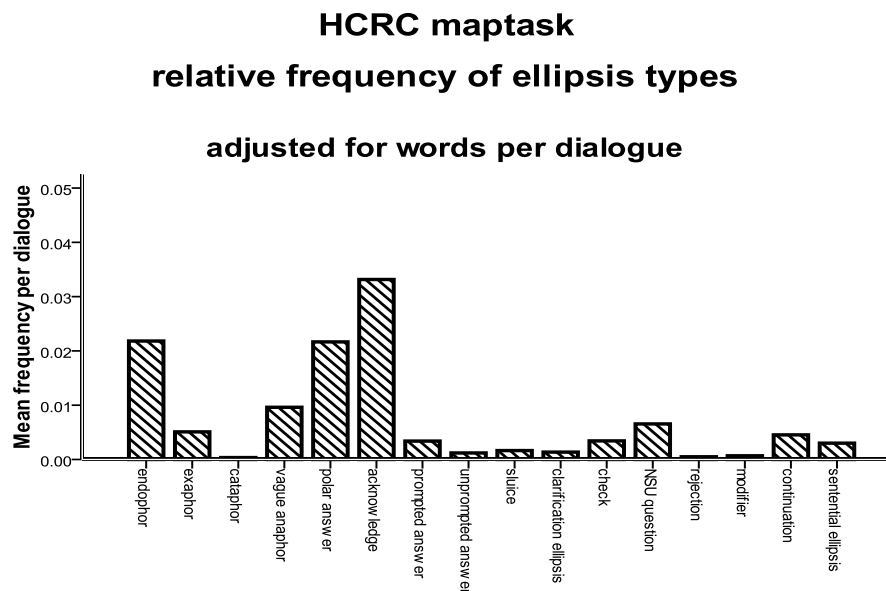


Figure 6: Frequency of ‘presumed shared context’/‘ellipsis’ categories in the HCRC Map Task corpus

In the map task dialogues there are higher rates of the NSU and sentential ellipsis types and a much lower use of anaphoric reference. In the HCRC map task it is the four types of NSU answer that are the most commonly used form of ellipsis, these are polar answers, acknowledgements, prompted and unprompted answers.

4.2.3 Medium and familiarity differences

Analysis of variance was carried out on the factors of medium and familiarity using frequency scores adjusted for words (all $df = 1, 124$). Medium differences were found in the following coding categories: combined giver and follower vague anaphor; combined giver and follower polar answers. Familiarity differences were found in the following categories: giver prompted NSU answers; giver sluice; giver NSU questions; follower NSU acknowledgements. No interaction effects were found between the experimental factors of medium and familiarity. The results are shown in Table 14.

Table 14: Medium and familiarity differences for the ‘presumed shared reference’ codings in the HCRC map task (ANOVA)

variable	mean	sd	mean	sd	F	sig
	Face-to-face		Audio only			
G and F V.A.	.0106	.0057	.0086	.0049	4.392	.038
G and F Polar answer.	.0202	.0069	.0232	.0080	5.178	.025
	familiar		unfamiliar			
Giver P.A.	.0044	.0032	.0033	.0030	3.952	.049
Giver sluice	.0005	.0008	.0002	.0007	4.312	.040
Giver NSUQ	.0017	.0018	.0010	.0016	5.343	.022
Follower Acknowledge	.0908	.0586	.1175	.0740	5.113	.025

From the above table it can be seen that with regard to medium, the category vague

anaphor was higher in the FTF condition; polar answers were higher in the audio condition. These differences only emerge with follower and giver role conditions combined; the effect of role is not strong enough to discriminate between them. For the familiarity factor, significant differences were higher in the familiar condition except for follower acknowledge. For these categories the effect of role discriminates between giver and follower.

4.2.4 Role differences

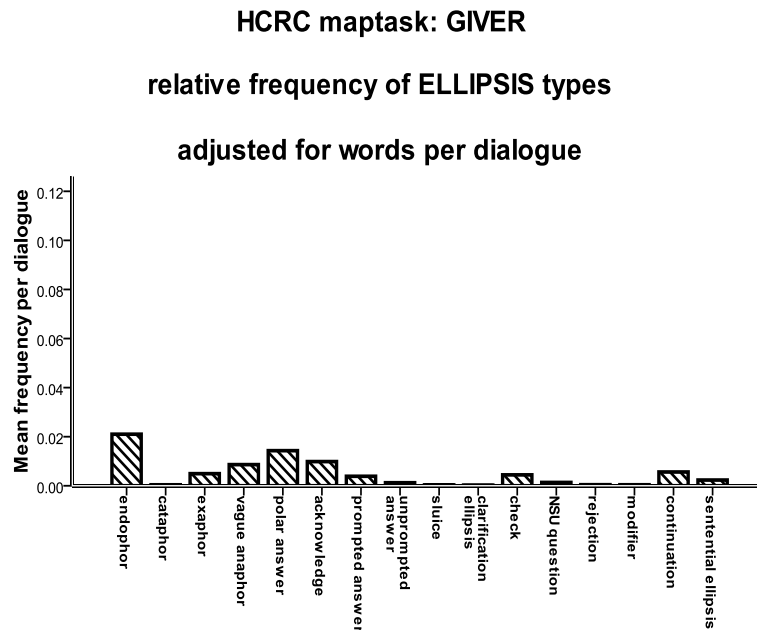


Figure 7: Frequency of ‘presumed shared context’/‘ellipsis’ categories in the ‘Giver’ role in the HCRC Map Task corpus

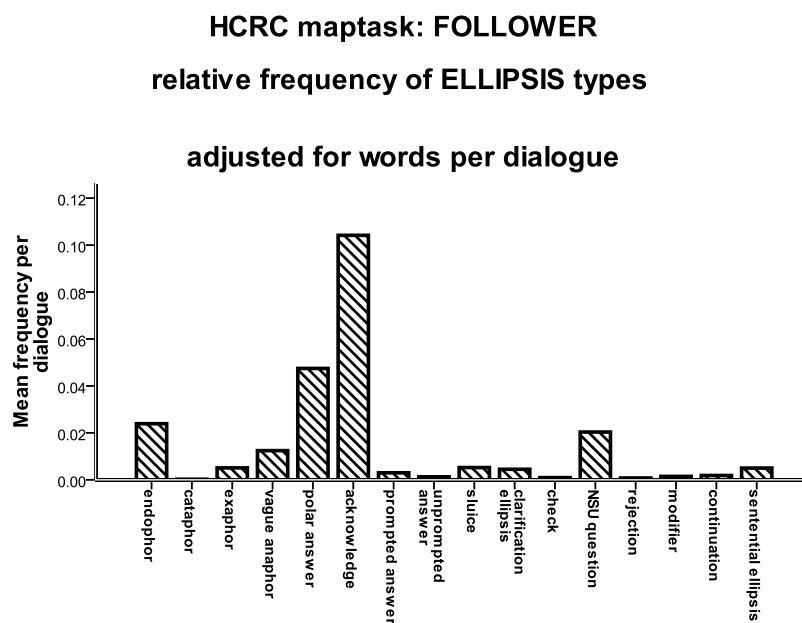


Figure 8: Frequency of ‘presumed shared context’/‘ellipsis’ categories in the ‘Follower’ role in the HCRC Map Task corpus

Figures 7 and 8 above illustrate visible differences in the profiles of ‘ellipsis’ or presumed shared context use for the two task roles using category frequencies adjusted for word totals. Table 15 below illustrates the significant differences (all $df = 1, 254$).

Table 15: ‘role’ differences in the ‘presumed shared reference’ coding of the HCRC map task (ANOVA)

variable	mean	sd	mean	sd	F	sig
	Giver		Follower			
V.A.	.0086	.0055	.0124	.0097	15.13	<.001
Polar answer	.0143	.0084	.0474	.0366	99.73	<.001
Acknowledge	.0099	.0073	.1042	.0678	244.72	<.001
Sluice	.0004	.0007	.0052	.0063	76.43	<.001
C.E.	.0003	.0007	.0045	.0067	51.70	<.001
Check	.0044	.0055	.0009	.0023	45.03	<.001
NSU Q	.0014	.0017	.0204	.0132	259.24	<.001
Modification	.0004	.0011	.0015	.0040	8.44	.004
Continuation	.0056	.0043	.0019	.0031	65.15	<.001
Sentential	.0023	.0024	.0050	.0067	18.06	<.001

Table 15 shows that there are a large number of statistically significant differences within the distribution of elliptical phenomena in the task roles; a simpler breakdown of the rate of ellipsis type by role is given below in Table 16.

Table 16: Rates of 'anaphora' and 'NSUs' by 'task role' in the HCRC map task

role	ellipsis	Sum total	words
Giver	Anaphora	3898	107233
Giver	NSU	4508	107233
Follower	Anaphora	2121	49082
Follower	NSU	7191	49082

Total number of turns containing NSUs for followers is 7191, out of a total of 9567 turns. It can be seen that 75% of follower turns are non-sentential (almost two-thirds of those being polar answers or acknowledgements). Task role or demands have a high impact on the use of non-sentential contributions to dialogue.

4.3 Correlational data

4.3.1 Correlations between coding categories

Tests were made on the correlation between the presumed shared context categories. In order to reduce the number of tests to be considered categories were grouped into anaphora, NSU answers, NSU questions, NSU statements (sentential ellipsis was another category). These five separate categories were then tested on three levels: at the level of dialogue (no distinction was made between task roles); within task roles (giver and follower intracorrelations); between task roles (giver and follower intercorrelations).

All significant correlations were in a positive direction; there were no negatively correlated categories.

- 1) Dialogue level analysis: there is a significant positive correlation ($P < .01$) between all of the five category groups.
- 2) Within-role intracorrelations: for followers there is a significant positive correlation ($P < .05$) between all of the five category groups. For givers there is a significant positive correlation ($P < .05$) in all but one pair; between anaphora and NSU statements (Spearman $\rho = .169$, $P = .056$).
- 3) Between-role intercorrelations: for follower-giver correlations all but three out of 25 were significant at the $P < .05$ level. The three correlations that were not significant were follower statements – giver sentential ellipsis ($\rho = .162$, $P = .068$),

follower sentential ellipsis – giver anaphora ($\rho = .169$, $P = .057$), follower sentential ellipsis – giver statements ($\rho = .052$, $P = .559$).

The individual values for the above results can be seen in tabulated form in Appendix F. These results are examined further in terms of predictive relationships in Chapter 6 using stepwise multiple regression.

4.3.2 Accuracy correlations

Correlations with deviation from the route were calculated using frequencies adjusted for words, shown in Table 17.

Table 17: Significant correlations of ‘presumed shared referent’ categories with deviation from the route (accuracy) in the HCRC map task

variable	Pearson’s r	P
G and F acknowledge	.216	.014
Giver continuation	.209	.018
Follower polar ans.	-.249	.005
Follower reject	.230	.009

NSU acknowledgements are not significant by role; only the combined giver and follower scores significantly correlate with accuracy. Follower polar answers are negatively correlated with deviation from the route; an increase indicates better accuracy. All acknowledgements, giver continuations and follower rejections have positive correlations with deviation (inaccuracy) and may be illustrating communication problems in the task.

4.4 Summary of results

Anaphoric reference is used more frequently in naturally occurring dialogue than task-oriented. Non-sentential contributions and sentential ellipsis are more frequent in the map task dialogues than the BNC. For the HCRC map task corpus in the face-to-face condition more vague anaphora were used but less polar answers. Familiarity of dialogue partner had a separate effect on the two roles with givers using more prompted answers, sluices and non-sentential questions and followers using more non-sentential acknowledgements. Task role has a noticeable effect on the frequency

of the ellipsis categories. Givers use more NSU checks and continuations; followers use more vague anaphora, polar answers, acknowledgements, sluices, clarification ellipsis, NSU questions, modifications and sentential ellipsis. This can be simplified as saying that givers use more anaphora; followers use more NSUs and sentential ellipsis. Multiple positive correlations were found between categories; using one type is associated with using other types. This is found both within and between task roles. Accuracy in the map task is found to be correlated with follower polar answers. Less accurate task routes are associated with acknowledgements, giver continuations and follower rejections.

4.5 Discussion

The results presented here have demonstrated a complex pattern of ellipsis use in the HCRC map task corpus, with differences to the pattern found in the BNC sample and across the experimental factors. Anaphora are a useful measure of mutual-understanding due to their nature, but as the map task involves discussion of negotiating one obstacle at a time there was little misunderstanding of pronoun use. This is not the case in naturally-occurring conversation; use of multiple anaphora with different antecedents can be found. Non-sentential utterances are found in both corpora. Due to the nature of the map task more concise, non-sentential contributions are made. When eye contact is not possible more polar answers are produced, presumably due to the use of head nods and shakes being used in the FTF condition. The higher use of vague anaphora in the FTF condition implies that cues are available visually that add meaning to the anaphor. Both of these findings support the ‘social cues’ hypothesis (as in media richness theory) that assumes that extra information is accessible when eye contact is possible, however this information does not affect the final accuracy score so may be of limited use depending upon the task or conversational goals. Unfamiliar pairs show higher follower acknowledgements; this may be due to a social need to demonstrate your ability to listen and understand when the other has no idea of your abilities. With familiar pairs the giver uses more prompted answers, sluices and NSU questions. Familiarity would seem to be associated with a brusque manner of questioning, for example using single wh- words to clarify or request further information. Givers checked that

they were being understood more and continued on from a previous utterance more; this is due to the nature of being an information giver. The categories that are higher in the follower condition are mostly questions, answers and acknowledgements, again unsurprisingly. The very high number of significant positive correlations was unexpected, but demonstrates that ellipsis use is associated between all categories. One ellipsis category was found to be positively correlated with accuracy, follower polar answers. This would seem to be indexing something more than the single category; a polar answer can only be given when a relatively specific question has been asked. This may be indexing to some extent the ability of the giver at the task, and the polar response is demonstrating a good degree of interactive coordination between the pair. Note that this is the only category from both the repair and ellipsis coding that positively indexes success at the task (some other categories are negatively correlated). Correlates and predictors of task accuracy are discussed further in Chapter 6.

The measures that have been presented here distinguish between naturally-occurring and task-oriented dialogue, medium, social norms (through familiarity) and individual goals (task role). From this it would seem that the ellipsis coding is a valid and useful tool for examining mutual understanding in dialogue.

4.6 Review

The coding scheme used here demonstrates a much higher use of elliptical contributions than previous studies; this is explained by the inclusion of anaphora. Naturally occurring dialogue and task-oriented dialogue show different patterns of ellipsis use. In a BNC sample anaphora were much more frequent but non-sentential utterances and sentential ellipsis were less frequent than in the map task. Experimental factors of medium and familiarity showed different distributions of elliptical categories over conditions. In the audio-only condition there were more NSU polar answers, acknowledgements and checks than in FTF. In the familiar condition there were more prompted answers, sluices and NSU questions than in the unfamiliar condition. The patterns of ellipsis use found in the HCRC map task show

the most marked difference between the task roles. Followers used more anaphora and more NSUs. 75% of followers' turns were non-sentential. Several categories are significantly correlated with accuracy in the map task.

Chapter 5

Negative indices of mutual understanding in the HCRC map task corpus

5.0 Overview

Healey et al. (2005) present a protocol for coding instances of repair phenomena in dialogue. A variation of this protocol created for this thesis identifies the seven most common repair types; this protocol is applied to the entire HCRC map task corpus. The results of coding a BNC sample for repair are contrasted with the repair frequencies in the map task. Factors of medium, familiarity and task role are examined; also the specificity of repair initiations.

5.1 Method

The method follows that outlined in the previous chapter; the coding protocol for repair and repair initiation was applied to a sample of transcripts from the BNC and transcripts of the entire HCRC map task corpus. These data were put into SPSS spreadsheets for analysis.

5.2 Results

Below are presented both a comparison of the BNC sample and the HCRC map task coding and a breakdown and analysis of the map task data. Firstly, the baseline measure of repair in the BNC sample is contrasted with the overall profile of repair found in the HCRC map task. Secondly the occurrences of repair in the map task are presented in detail. Differences across the factors of medium, familiarity and role are given, then a breakdown of the specificity level of repair initiations (P2NTRI).

5.2.1 British National Corpus vs. HCRC map task data

Figures 9 and 10 below show that a different pattern of repair emerges between the map task and BNC coding (exact figures are given in Table 18 below). This clearly shows that the patterns of distribution are different in the two corpora; more repair is

found in the task-oriented dialogue. Only position one transition space repairs are more frequent in the BNC sample of naturally-occurring dialogue.

Table 18: Relative frequency of ‘repair’ categories in the HCRC map task and a BNC sample

	P1A	P1F	P1TS	P2OIOR	P2NTRI	P3OISR	P3SISR
BNC	.0052	.0130	.0023	.0006	.0026	.0016	.0032
Map Task	.0117	.0193	.0004	.0028	.0074	.0049	.0032

The figures in Table 18 are a ratio score of the frequency of repair types divided by the number of words in the dialogue. Note that the BNC baseline figures are intended to give an approximate illustration of the naturally occurring distribution; a larger sample would ideally be coded for a more complete picture. Note also that position one articulation repairs especially are susceptible to differences in transcribing; in the BNC the dialogue is sometimes rather tidied-up with regard to ‘disfluent’ speech. For these types of repair the CA style of transcription that includes all produced sounds would be most appropriate for coding; this is not easily done for experimental studies due to its extremely time-consuming nature.

Due to the significantly higher number of turns found within the audio-only and familiar-pairs conditions, simple frequency counts of repair types were thought to be misleading as the frequency totals reflect the length of the dialogue (ie. more talk = more repairs). As the cleaned-up format of the available transcripts gave equal numbers of turns for both givers and followers, it was decided that word counts for givers and followers from each dialogue should be used to normalise repair frequencies, removing a potentially confounding effect of dialogue length, and are used in the following analyses.

5.2.2 Repair Frequencies

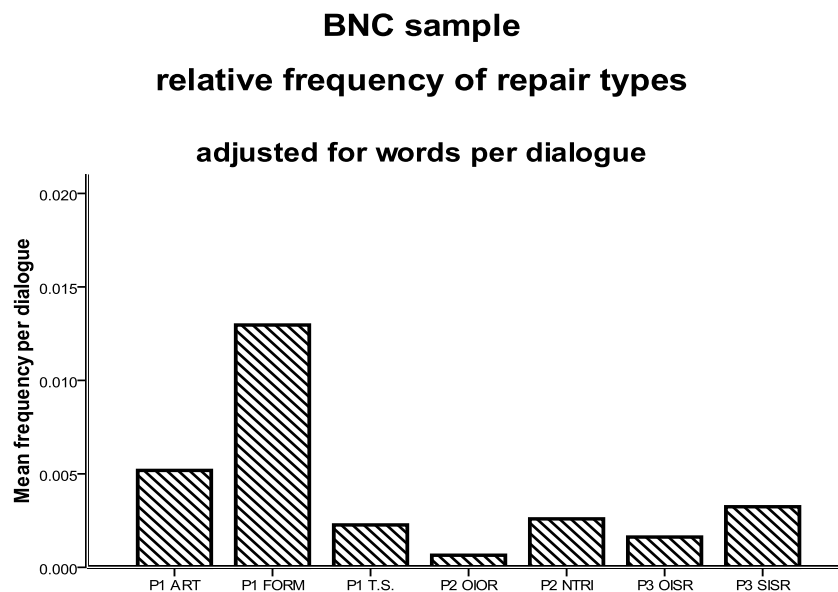


Figure 9: Frequency of ‘repair and repair initiation’ categories in a sample from the BNC

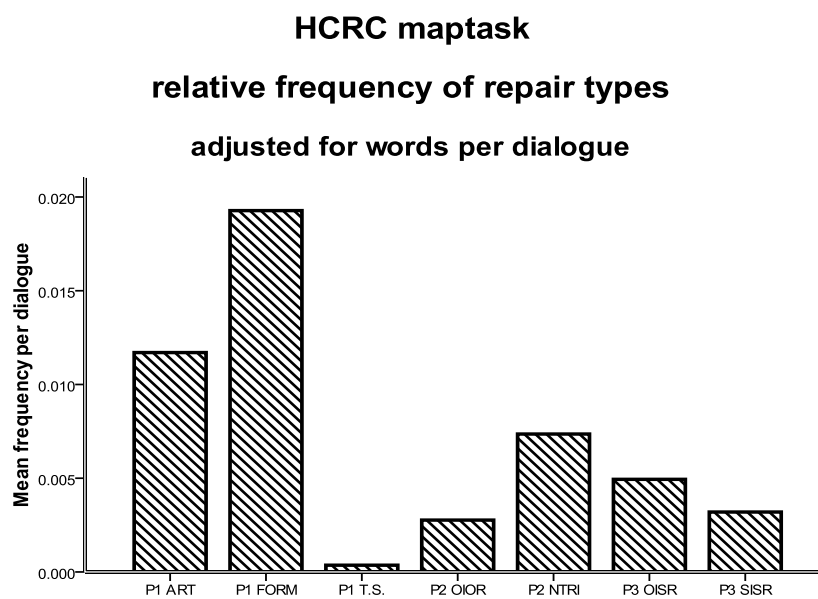


Figure 10: Frequency of ‘repair and repair initiation’ categories in the HCRC Map Task corpus

5.2.3 Medium, familiarity and role differences

The HCRC map task corpus data was tested for differences across the conditions of medium, familiarity and task role.

Table 19: ‘Repair’ differences in audio versus face-to-face conditions in the HCRC map task using standardised data (ANOVA)

variable	mean	sd	mean	sd	F	sig
	Face-to-face		Audio only			
Follower P3 SI SR	.0019	.0032	.0030	.0031	3.996	.048

There is a tendency for followers to use more P3 SI SR in the audio only condition; all other repair categories are similar across the medium conditions. When tested using unadjusted frequencies, one further difference is significant; giver position one transition space repairs are more frequent in the face-to-face condition (Mann-Whitney U = 1735; p = .018).

An ANOVA test (all df = 1, 124) showed that there were no significant familiarity effects and no interaction of familiarity and medium. Table 20 below shows the single repair category that was approaching significance over the familiarity conditions.

Table 20: ‘Repair’ differences in familiar versus unfamiliar conditions in the HCRC map task using standardised data (ANOVA)

variable	mean	sd	mean	sd	F	sig
	Familiar		Unfamiliar			
P2NTRI	.0080	.0036	.0067	.0043	3.66	.058

Examination of specificity levels using adjusted repair/words ratios does not identify which types of P2NTRI are significantly affected by familiarity. When words are not accounted for, ie. raw frequencies, it is the highest two specificity levels of follower P2NTRI that are significant. Nonparametric Mann-Whitney tests were used, shown in Table 21.

Table 21: ‘P2NTRI specificity level’ differences in familiar versus unfamiliar conditions in the HCRC map task (Mann-Whitney tests)

variable	mean	sd	mean	sd	U	sig
	Familiar		Unfamiliar			
Follower spec 4	3.31	2.87	2.33	2.43	1599.5	.030
Follower spec 5	3.39	2.84	2.47	2.48	1613.5	.036

The followers in the familiar-pairs condition used more of the highest specificity repair initiations (‘You mean...?’).

There are differences in the repair profiles of the two roles in the map task, here illustrated in Figures 11 and 12.

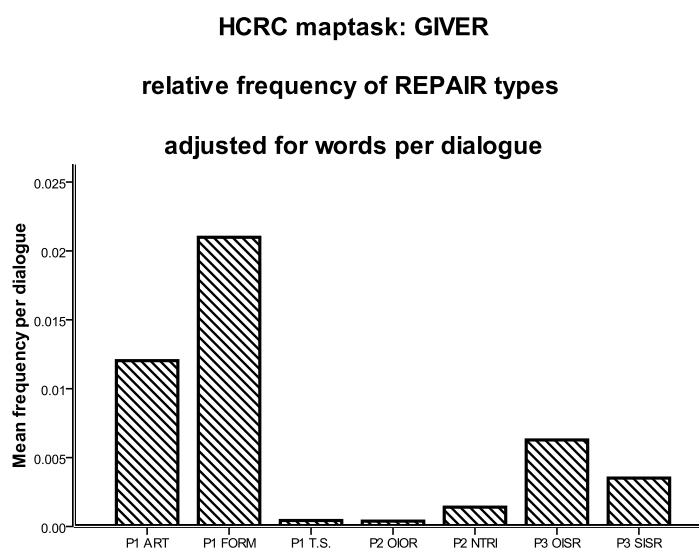


Figure 11: Frequency of ‘repair and repair initiation’ categories in the ‘Giver’ role in the HCRC Map Task corpus

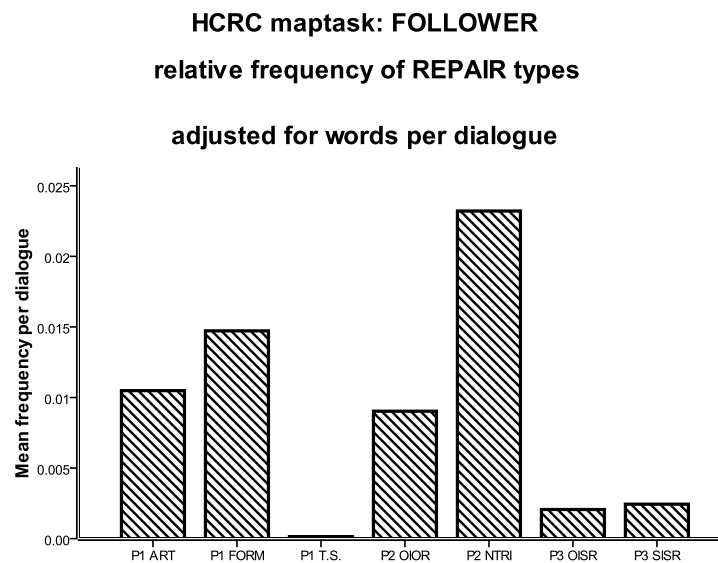


Figure 12: Frequency of ‘repair and repair initiation’ categories in the ‘Follower’ role in the HCRC Map Task corpus

Of the seven repair types identified in the protocol, only two types are not significantly different; position one articulation (P1A) and position one transition space repair (P1TS). Of the remaining repair types, in two cases the follower's mean frequency is higher; position two other-initiated other repair (P2OIOR) and position two next turn repair initiation (P2NTRI). In the remaining three cases the giver's mean frequency is significantly higher; position one formulation repair (P1FORM), position three other-initiated self repair (P3OISR) and position three self-initiated self repair (P3SISR). For role differences all $df = 1, 254$.

Table 22: 'repair' category differences in 'role' conditions in the HCRC map task (ANOVA)

repair	giver		follower		F	p
	mean	sd	mean	sd		
P1art	.0120	.0079	.0105	.0085	2.29	.132
P1form	.0210	.0098	.0147	.0088	28.87	.000
P1 TS	.0004	.0016	.0001	.0007	3.50	.063
P2 OIOR	.0004	.0007	.0090	.0086	128.34	.000
P2NTRI	.0014	.0020	.0234	.0157	247.13	.000
P3 OISR	.0063	.0043	.0021	.0040	65.93	.000
P3 SISR	.0035	.0029	.0024	.0032	7.86	.005

5.2.4 Specificity of next-turn repair initiations

It has been shown that followers use more position two repair initiations (P2NTRI) than givers; here that difference is expanded to illustrate the levels of specificity used by the two respective task roles. From Figures 13 and 14 below it can be seen that there is a preference for the strongest, or most explicit, forms of P2NTRI. Raw frequencies have been used in order to further illustrate the effect of role on overall usage.

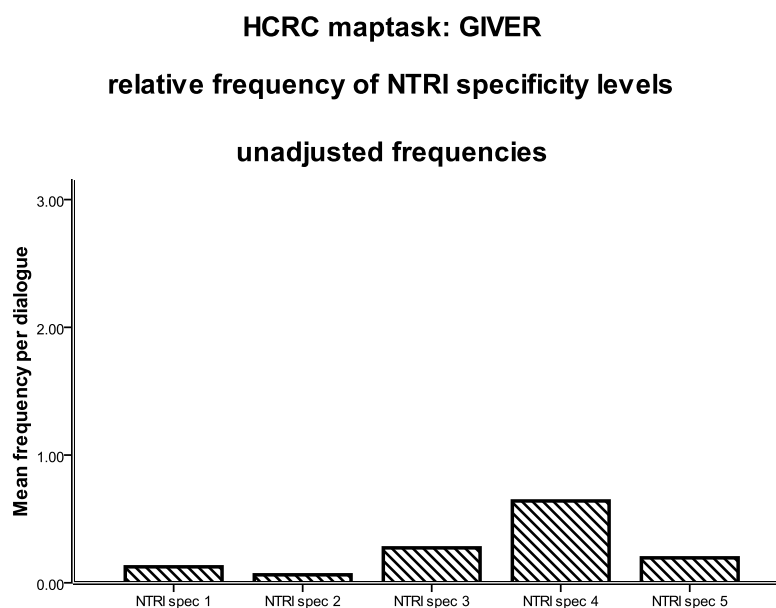


Figure 13: Frequency of ‘Next-Turn Repair Initiation’ specificity levels 1-5 in the ‘Giver’ role in the HCRC Map Task corpus

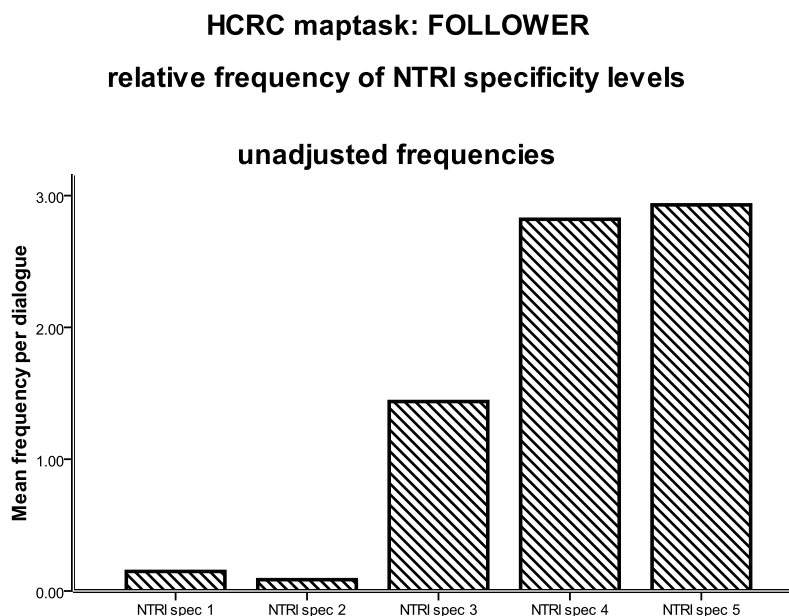


Figure 14: Frequency of ‘Next-Turn Repair Initiation’ specificity levels 1-5 in the ‘Follower’ role in the HCRC Map Task corpus

It was thought that a predictive relationship may emerge between accuracy in the task and the specificity level of P2NTRIs; higher specificity levels indexing a higher level of coordination and mutual understanding. However, regression of the P2NTRI specificity levels onto accuracy did not produce a reliable model; investigation showed that route accuracy did not significantly correlate with any of the five levels. Further regression analyses combining the repair and ‘ellipsis’ codings are presented and discussed in Chapter 6.

Follower P2NTRI specificity levels were tested for intercorrelations (giver P2NTRIs were too infrequent for valid analysis). It was postulated that a negative relationship may emerge between the high and low levels, i.e. highly specific requests for clarification would be associated with fewer of low specificity (e.g. ‘huh?’). For these data this is not the case; the significant relationships are all positive.

Table 23: ‘P2NTRI’ repair category correlations within the ‘follower’ role in the HCRC map task

P2NTRI specificity levels		Spearman Rho	p
Spec 1	Spec 5	.205	.021
Spec 3	Spec 4	.275	.002
Spec 4	Spec 5	.347	<.001

5.3 Relationships between variables

A further exploration of the data is possible by examining the correlations between repair phenomena. Nonparametric Spearman correlations are used as in Table 23 above; this has the advantage of having no assumptions about the data. For these analyses the number of words produced is irrelevant, only the frequency of repair phenomena are being examined. Table 24. presents correlates of the different follower P2NTRI specificity levels. Table 25 presents a wider examination of the relationship between the repair categories, both between the giver and the follower and within the task role.

Table 24: ‘Repair’ categories correlating with different ‘follower NTRI’ specificity levels (1=lowest; 5=highest) in the HCRC map task

NTRI specificity level	Correlating variable	Spearman Rho	P
1	GP2 NTRI	.221	.012
1	GP3 OISR	.204	.021
3	GP3 OISR	.502	<.001
4	GP1 Art.	.293	.001
4	GP1 Form.	.309	<.001
4	GP2 OIOR	.216	.014
4	GP3 OISR	.683	<.001
4	GP3 SISR	.290	.001
5	GP1 Art.	.259	.003
5	GP1 Form.	.188	.033
5	GP2 NTRI	.178	.045
5	GP3 OISR	.599	<.001
5	GP3 SISR	.175	.048

Table 25: Correlational relationships between ‘repair’ categories in both ‘role’ conditions in the HCRC map task

repair		Spearman Rho	p
Variable 1	Variable 2		
Between roles			
GP1 Form.	FP1 Form.	.293	.001
GP1 T.S.	FP1 T.S.	.310	.000
GP2 OIOR	FP1 Art.	.189	.033
GP2 OIOR	FP3 OISR	.308	.000
GP3 OISR	FP2 OIOR	.214	.015
FP2 NTRI	GP2 OIOR	-.261	.003
FP2 NTRI	GP2 NTRI	-.266	.002
Within roles			
GP1 Form	GP2 OIOR	.187	.034
GP2 NTRI	GP2 OIOR	.343	.000
FP1 Art.	FP3 SISR	.248	.005
FP1 Form.	FP3 SISR	.178	.044

From Table 24 it can be seen that all but one (level 2) of the ‘follower’ NTRI specificity levels significantly correlates with the respective ‘giver’ P3 OISR.; when a problem is flagged the originator addresses it. Figures 12 and 13 above show that NTRI specificity level 2 is the most infrequent of initiation types; it would not be expected to correlate strongly with any categories. NTRI specificity levels 4 and 5

are related to giver position one self-repair of both types.

5.4 Summary of results

As discussed previously, the CA tradition strictly avoids quantification of conversational phenomena. The study presented here suggests that the identification of repair types can be used to provide a quantitative index of interaction; this can be used both to discriminate between conditions and examine the relationship between different repairs within an interaction. In the current study this can be most clearly seen in the differences between the task role performed (either instruction giver or instruction follower). For the task role, very different patterns of repair type were used; for the giver, it was much more common to repair in either position one or position three. In contrast, the follower used position two repairs and repair initiations more frequently. Repair initiations were preferred to be explicit, but use of highly specific initiations was associated with more general trouble signalling. Medium has an effect on one type of self-repair in position three, usually a response to a response. Familiarity appears to have an effect on the type of trouble signal used, but in these data the effect is not strong. Overall, the pattern of repair types used differed from that found in a sample of naturally-occurring dialogue.

5.5 Discussion

One aspect of this study was intended to show that analysis of process measures of interactions may show significant differences between conditions that do not emerge from standard outcome measures. In the case of the HCRC map task, it can be seen that accuracy as measured by deviation from the given route does not discriminate between media; however, some differences have been found in relation to trouble signalling and repair for medium, familiarity and task role conditions. For example, using unadjusted frequency scores it was found that giver transition space repairs occurred more frequently in the eye contact condition; this may be due to the giver being aware of non-verbal signals that allow them to judge whether they are making themselves clear or not, and adjusting their contribution accordingly. Conversely, lacking this ability led to an increase of followers' position three self-initiated self repairs, which were more common in the no eye contact condition (this result found

using both adjusted and unadjusted frequencies). In the case of position two next turn repair initiations being more common in the familiar pair condition, it may be that participants felt more comfortable questioning the other's contribution when they already knew them.

The quantitative approach presented here allows the relationship between different repair phenomena frequencies to be examined. In two cases it can be seen that giver position 1 articulation phenomena are positively correlated with follower position two NTRI (specificity levels 4 and 5); this may show that if someone is struggling to produce their contribution there is a greater likelihood that it will be treated as a trouble source. This occurs when the exact problem area is identifiable; there is no significant correlation with the less specific NTRI types and demonstrates that articulation self-repairs are highly localised within a contribution and do not interfere with the overall comprehension. Further significant correlations were found, such as the positive relationship between both givers' and followers' formulation self-repairs. It may be that this problem with formulating a contribution lessened over time for both roles and was connected to familiarity with the task; further studies would be needed to test this. Negative relationships were found between follower NTRIs and givers' position two repairs and initiations. Followers are attempting to carefully monitor the instructions fed to them by the giver; it may be that more follower NTRIs indicate a higher level of monitoring and this is reflected in a more precise contribution style that does not create trouble and hence less giver NTRIs and corrections. However, despite this follower NTRI does not significantly correlate with accuracy; the overall influences on task success are complex. Within the task roles there were positive correlations between follower position one and position three self-initiated self repairs. This is not mirrored in the giver condition and may demonstrate that the role of follower requires more self-monitoring and a precise manner of conveying meaning.

The results from the repair data study suggest that the role within the task being performed by participants had a significant effect on the frequency of different repair types, at variance with the repair profiles found in naturally occurring dialogue.

Among the findings were that givers used significantly more of all self-repair types, while followers used more repair initiations and other-repairs. Familiarity was shown to have a significant effect on the use of NTRIs, with familiar followers using more. Eye contact was found to have a significant effect on the use of self-repair; givers in the no eye contact condition used more self-repairs in position three when there had not been an NTRI in position two. Other types of conversational repair are shown to correlate with each other, distinguishing between both medium and familiarity of participants. These results suggest that quantitative analysis of repair may be a useful tool for analysing various measures of communication, and understanding how mutual-understanding is achieved.

5.6 Review

The analysis of the HCRC map task shows few outcome differences over conditions. Both the audio-only and familiar pair conditions contained more turns; there was no interaction. Accuracy in the task was not different over conditions. Repair in the map task demonstrates a different profile to that found in the BNC sample; in all categories except position one transition space more repair was found in the map task. The ratio of repair to words was much higher in the map task. The major medium difference was more follower P2 NTRI in the audio-only condition; the familiar condition also showed a higher rate of follower P2 NTRI. Repair profiles for the two roles were very different; givers used more position one and three repairs, followers used more position two repairs and initiations. The number of giver words positively correlated with accuracy. Repairs of all types from the protocol were found in the corpus. Followers produced more NTRIs and demonstrated a preference for high specificity. P2 NTRIs negatively correlated with P2 OIORS, with initiations preferred over repairs.

Chapter 6

The relationship between positive and negative indices of mutual understanding

6.0 Overview

In this chapter the results from the repair and ellipsis studies will be integrated in order to identify dimensions indicating task success and the overall influence of medium, familiarity and role. These results can then be seen in the wider context of theories of communication and media. The codings of the map task using DGA (Carletta et al., 1996) were also included in some analyses due to potential similarities in some of the categories; this allowed comparison of the different approaches to categorising the dialogues.

It has been shown in the previous two chapters that naturally occurring dialogue measured by a sample from the BNC demonstrates different distributions of both repair and ellipsis contrasted with the distributions found within the map task data.

In only one type of repair does the BNC sample show a higher frequency than the task-oriented dialogue; position one transition space repair. Even the highest occurring condition (given FTF P1 TS) did not approach the frequency found in the BNC sample; .0006 contrasted with .0023 (mean ratio of frequency to words). In all other repair categories more repairs are found in task-oriented dialogue. Although both the BNC sample and the map task corpus demonstrate the preference for self-correction (Schegloff et al., 1977) in the task-oriented dialogue there are over four times as many position two other-initiated other-repair (P2OIOR) than in the BNC sample. There are over twice as many position one articulation repairs in the task-oriented data; other differences are not as noticeable although higher than in the BNC sample.

The use of anaphora is much more prevalent in the BNC sample. Polar answers and acknowledgements are more prevalent in task-oriented dialogue; other NSU or ellipsis types are of relatively low frequency in both cases.

These data are useful in that they demonstrate a measurable difference between naturally occurring dialogue and task-oriented dialogue. Further examination of these data will allow some tentative conclusions to be drawn regarding the process of communication when attempting a task.

6.1 Comparison of results from all approaches

Analysis of a sample of dialogue from the BNC demonstrated the relationship between certain ellipsis and repair categories. For simplicity the 'ellipsis' categories were condensed into anaphora, NSU questions, NSU answers and NSU statements plus sentential ellipsis. All repair types were correlated with all ellipsis types (Spearman $\rho = .401$; $p < .001$). Anaphora is correlated with all P1 repairs ($\rho = .408$; $p < .001$). NSU questions correlate with both NTRIs ($\rho = .453$; $p < .001$) and P3OISR ($\rho = .180$; $p = .023$) i.e. the clarification request-response sequence. P3OISR is also correlated with NSU answers ($\rho = .161$; $p = .042$). These data show that in general the more repair there is, the more anaphora and ellipsis also. This relationship still holds when using ratios of phenomena/words rather than raw frequencies (all ellipsis-all repair correlation; Pearson $r = .531$; $p < .001$).

One of the aims of this research was to examine if there was a predictable relationship between aspects of the process of dialogue and performance accuracy in the map task. Preliminary analyses of correlations suggested that in both of the approaches described (presumed shared context; repair) there were useful associations with task performance, shown below. Significant DGA correlations are also included for comparison and contrast.

Parametric correlation coefficients are calculated using ratio data (standardised for words) as this should give a more accurate picture than nonparametric tests. These are shown in Table 26 below.

Table 26: Categories from ‘repair’, ‘presumed shared referents’ and ‘DGA’ coding schemes that significantly correlate with ‘deviation from the route’ using standardised scores

coding	type	Pearson’s r	p
Repair	Follower P1 Formulation	.170	.055
ellipsis	Giver continuation	.209	.018
	Follower polar answer	-.249	.005
	Follower rejection	.230	.009
DGA	Giver instruct	.360	.001
	Follower instruct	.248	.008
	Follower reply Y	-.248	.008
	Follower reply N	-.203	.032

6.1.1 Accuracy and instructions

The results in Table 26 illustrate the variables that positively correlate with deviation; as these increase the drawn route becomes less accurate. These include ‘giver instruct’, one of the most common DGA categories and essential for the task. It is possible that too many instructions from the giver confuse or bore the follower to such an extent that the route suffers. However, follower instructions are also associated with a less accurate route. Instructions appear to be a bad thing with regard to task success. This may be connected to psychological effects of instructions from an experimenter reducing reported interest in a task (e.g. Sansone et al., 1989) although the effect here seems to be broader in that instructions from either of the participants during the task is associated with lower task accuracy. It may be that when people are struggling to understand one another they produce and respond to more instructions but these extra turns do not, beyond a certain point, actually improve performance.

6.1.2 Rejections and continuations

NSU rejections and continuations occur relatively rarely; the fact that they have a significant relationship to a less accurate or successful task suggests that despite their infrequent usage they may be a useful index of task success. These are the definitions from the ellipsis protocol:

Rejection: A sentence fragment is used to dispute, reject or correct a previous statement.

Continuation: a sentence fragment that directly continues from a previous contribution.

Rejections are positively correlated with deviation from the route; this category is not capturing the same phenomena as P2 OIOR repairs despite including corrections (Pearson $r = .116$; $p = .193$).

For example:

(q5ec2)

Giver: oh there's another plane crash down here though

Follower: **no only one**

Rejection phenomena are associated with a worse route and may index a lack of understanding of the task or the forming of false assumptions; in the above example one assumes that the other has exactly the same landmarks. It may be that rejections are indexing a lack of mutual context; a specific problem in common ground has emerged. In this case it may be that these phenomena are illustrating a general aspect of common ground, rather than being specific to the task. If each person presents their (incompatible) version of some phenomena presumed to be grounded a 'rejection' emerges.

Continuations are also associated with a less accurate route; this may be due to the confounding effect of the two possible different types:

- i. The listener continues the contribution on from the originator.

ii. The original speaker has paused or been interrupted and then continues. In the same manner as for repair, a continuation may be carried out by ‘self’ or ‘other’. In the case of (i) a higher amount of coordination is required than (ii); in the latter case the speaker may be ignoring the listener. If the case (ii) is more prevalent this category would be indexing poor coordination; this may be the case here. This phenomenon is not related to P1 Transition Space repair (Pearson $r = .031$; $p = .73$).

The below examples have surface similarities to third turn repair, but it can be seen that this is not repair; the forward momentum has not stopped due to trouble.

(1) Q2EC5

Follower: I’ve got flamingoes down here with a
 Giver: yeah.
 Follower **quite a small gappy bit.**

(2) Q1NC6

Giver: Yeah, it’s about ... you know an inch or so two inches from ...
 to the left of the old pine, so if you just take a line straight
 down from the start, okay?
 Follower: uh-huh, so ju—hang on you see this ehm?
 Giver: **four inches down... four inches down**

(3) Q5NC5

Follower: so I’m making a curve round the diamond mine yeah?
 Giver: no no no no
 Follower: no no no no
 Giver: f-- ... **past the diamond mine.**

6.1.3 Formulation repairs

Follower position one formulation repairs correlate positively with deviation; it is notable that this is the only type of repair that correlates (either positively or

negatively) with deviation from the route. It may be that the reformulation of their contribution is due to some aspect being perceived as unclear in some way. This impinges upon the utterance to such an extent that the speaker alters the content of their contribution; Follower P1 formulation repairs give an index of perceived trouble and less successful communication. If the follower is having trouble the route drawn will suffer; the instructions from the giver do not affect it in this way.

6.1.4 Polar answers associated with accuracy

All NSU polar answers are negatively correlated with deviation from the route; a more accurate result. NSU polar answers and DGA reply-Y are approximately measuring the same thing (Pearson $r = .969$; $p < .001$) and both signify positive understanding of some point. This relationship also applies to DGA reply-N (correlation with NSU polar answers = $.788$; $p < .001$). Yes/no answers indicate that relatively precise unambiguous questions can be asked and this lack of potential confusion (relative to open-ended questions) is associated with a more accurate route.

6.1.5 Grounding and acknowledgements

The association of less accurate routes with NSU acknowledgements is not due to a simple instruction – acknowledgement relationship, where instructions are associated with lower accuracy. Giver instructions are correlated with follower NSU acknowledgements ($r = .348$; $p < .001$), but follower instructions are not correlated with giver NSU acknowledgements ($r = .105$; $p = .269$). Follower NSU acknowledgements alone are not correlated with accuracy ($r = .086$; $p = .332$).

Two types of acknowledgement are included in the analysis; from the ellipsis protocol (significantly associated with lower accuracy) and the DGA coding (not associated with accuracy). Clark and Brennan (1991) suggest that acknowledgements are one of the major means of demonstrating positive evidence of grounding; here this does not seem to be the case. In these data there is a significant positive relationship between NSU acknowledgements and deviation, or a less accurate route. Why might this be? It might be useful to see which other variables correlate with acknowledgements. Firstly, there is no significant correlation between the DGA

category ‘acknowledge’ and the NSU/ellipsis category. The differences are presumably due to the specific definitions used:

From the DGA coding scheme (Carletta et al., 1996):

[initiation, response or preparation?] ->

response ->

[does the response contribute task/domain information or does it only show evidence that communication has been successful?] ->

communication ->

[acknowledgement]

This can be contrasted with the ellipsis protocol term ‘acknowledgement’¹⁹:

[Does all or part of the contribution answer a question or query?] ->

Yes ->

[Does this contribution contain a sentence fragment, word or noise that acknowledges understanding or agreement with a previous contribution?] ->

Yes ->

[acknowledgement]

It seems likely that the ellipsis/NSU coding covers a wider range of backchannel signals and these obscure the DGA type evidence which Clark and Brennan refer to. Variables that most strongly significantly correlate (negatively) with these²⁰ acknowledgements are ‘disfluency’²¹ ($r = -.565$, $p < .001$) and all repairs ($r = -.269$, $p = .002$). Also all DGA categories negatively correlate significantly. This suggests that the positive backchannel signals being captured are signifying a lack of negotiation of understanding; apparently positive backchannels are being produced regardless of understanding. It should be noted that even if the DGA disfluency types are capturing the phenomena that Clark and Brennan propose are evidence of grounding, they are not associated with accuracy or success in the task. DGA disfluencies are correlated with all position one repair types ($r = .977$; $p < .001$), although position one articulation and formulation one repairs are measuring separate phenomena (discussed in section 3.4.6).

¹⁹ Note that polar answers have already been captured through the protocol.

²⁰ i.e. from the ellipsis protocol.

²¹ Coded by the HCRC group alongside DGA

6.1.6 Further evidence for grounding

Two further types of evidence for grounding (Clark and Brennan, 1991) are ‘relevant next turn’ (Sacks et al. 1974) and continued attention. How would these be identified within the coding schemes? ‘Continued attention’ does not seem to be useful here; in all cases the two participants listened to each other and completed the task, with varying levels of success. Relevant next turns can be examined; for example a question and answer form an adjacency pair. If an appropriate answer is given it is safe to assume that the question was understood at some level, if an inappropriate answer or other contribution was given this would be rectified by an NTRI from the question asker. However, neither giver nor follower NTRIs correlate with deviation, at all specificity levels.

6.1.7 Correlation and regression

Correlation tests illustrate a relationship or association between two variables, with no cause and effect implied. A similar approach to data is regression analysis, whereby predictors of a variable are calculated. Although more usually performed with an independent variable as the target, it is also possible to find predictors of a separate dependent variable which is the approach taken here (Howell, 1994). The following sections outline various analyses which investigate predictors of deviation (a measure of accuracy and possibly successful communication) as well as other relationships between variables.

6.2 Multiple Regression Analysis

The analysis of the HCRC map task data involves a large number of variables; in order to examine whether there are repair or ellipsis types that are particularly associated with accuracy in the task (measured as deviation from the route) a stepwise multiple regression analysis of all repair and ellipsis variables was carried out, firstly with frequency scores and secondly with frequency: words ratio scores. Both types of data are used as both can be considered valid; both raw frequencies and adjusted frequencies are necessary to give an overall picture of these data.

Table 27: stepwise multiple regression of predictors of ‘deviation from the route’ using raw frequencies from ‘repair’ and ‘presumed shared reference’ categories (only significant predictors are included)

variable	Multiple <i>R</i>	<i>B</i>	Standard error <i>b</i>	Beta	<i>t</i>	Significance of <i>t</i>
Follower NSU Polar answer	0.35	-1.99	0.50	-0.47	-3.99	.001
Follower NSU Rejection	0.49	20.80	6.74	0.37	3.09	.01

For the raw frequencies in the stepwise multiple regression, ‘follower polar answers’ was entered first and explained 12.3% of the variance in deviation from the route ($F_{1,62} = 8.721$, $P = 0.004$). ‘Follower NSU rejection’ was entered second and explained a further 12.1% ($F_{1,61} = 9.717$, $p < 0.001$). A more accurate route drawn on the map by the follower was associated with more follower polar answers but fewer follower NSU rejections. A more complex picture emerges using standardised scores; here it was also possible to include standardised DGA data categories.

Table 28: stepwise multiple regression of predictors of ‘deviation from the route’ using standardised data from ‘repair’, ‘presumed shared reference’ and ‘DGA’ categories (only significant predictors are included)

variable	Multiple <i>R</i>	<i>B</i>	Standard error <i>b</i>	Beta	<i>t</i>	Significance of <i>t</i>
Giver instruct	0.36	1288.16	284.52	0.358	4.53	.001
Follower reject	0.43	6714.55	1951.86	0.274	3.44	.001
Follower instruct	0.49	1122.64	381.82	0.233	2.94	.004
Follower P3 SI SR	0.53	-3853.86	1228.97	-0.253	-3.14	.002
Follower P1 Form	0.57	1275.85	459.47	0.223	2.78	.007
Giver NSUQ	0.60	5468.99	2363.70	0.183	2.31	.023

The results here show that the ‘giver instruct’ category from DGA was entered first and accounted for 12.9% of the variance in deviation from the route ($F_{1,110} = 16.35$; $P < .001$). ‘Follower rejection’ from the ellipsis coding accounts for a further 5.8% ($F_{1,109} = 7.71$; $P = .006$). ‘Follower instruct’ (DGA) accounts for a further 4.9% ($F_{1,$

$F_{1, 108} = 6.90$; $P = .01$); ‘Follower P3 SI SR’ (repair) for a further 4.1% ($F_{1, 107} = 6.18$; $P = .015$); ‘Follower P1 formulation’ for a further 4.7% ($F_{1, 106} = 7.39$; $P = .008$); ‘Giver NSU question’ (ellipsis) a further 3.3% ($F_{1, 105} = 5.35$; $P = .023$). In this analysis a more accurate route was associated with fewer giver instructions, follower rejections, follower instructions, follower position 1 formulation repairs, giver non-sentential questions but with more follower position 3 self-initiated repairs.

These data are correlational in nature but do suggest what the possible underlying mechanisms relating to successful task performance achieved through dialogue may be. In this case it seems that excessive instruction giving (by both givers and followers) indicated a worse route drawn. It may be that over a certain level giving instructions becomes an index of how poorly two people are connecting; instructions have to be repeated or elaborated due to low mutual-understanding. Follower ‘rejections’ occur when something contributed by the giver is disputed; it is likely that this effect is an artefact of the task, in that the giver may assume that the follower’s map includes the same landmarks as the giver’s. Follower position 1 formulation repairs index a difficulty with the content of a contribution; it appears that if the follower is unclear or uncertain on what to ask or tell, the route suffers. Giver non-sentential questions appear again to be an artefact of the task type; a giver NSU would typically follow a follower utterance that is not totally understood. Only one category is negatively correlated with deviation from the route (i.e. increases with accuracy); follower position 3 self-initiated self-repair. This occurs when a follower decides to change in some way a previous contribution of theirs. Visual inspection of transcripts suggests that these occurrences are corrections of statements about landmarks; typified by ‘oh, sorry, I do have that landmark’ statements. Although these may seem to index poor attention to the task, these contributions show sufficient vigilance to actually be a predictor of accuracy or task success; the absence of this type of repair means that mistakes go uncorrected.

6.2.1 Further regression tests

The preceding regression tests explored predictors of deviation from the route in the

map task. The regression tests below examine relationships within the data by examining predictors of other variables; high-specificity NTRIs, all ellipsis, all repairs and all anaphora.

Table 29: Stepwise multiple regression of predictors of ‘high specificity NTRIs’ from ‘repair’ and ‘presumed shared reference’ categories (only significant predictors are included)

variable	Multiple <i>R</i>	<i>B</i>	Standard error <i>b</i>	Beta	<i>t</i>	Significance. of <i>t</i>
P3OISR	0.41	0.88	0.15	0.43	5.86	.001
P2OIOR	0.48	-2.56	0.88	-0.23	-2.91	.004
G Sent	0.54	0.88	0.27	0.24	3.22	.002
G None	0.56	-0.18	0.06	-0.23	-3.03	.003
G VA	0.59	-0.33	0.13	-0.21	-2.59	.011

Giver P3 OISR was entered first and accounted for 16.6% of the variance ($F_{1, 126} = 25.04$; $P < .001$). Then, factors entered were respectively Giver P2 OIOR (6.1%; $F_{1, 125} = 9.79$, $P = .002$), Giver Sentential ellipsis (6%; $F_{1, 124} = 10.36$, $P = .002$), Giver ‘none’ ie. no ellipsis (3.1%; $F_{1, 123} = 5.66$, $P = .019$); Giver vague anaphor (3.6%; $F_{1, 122} = 6.71$, $P = .011$).

That GP3OISR should be the main predictor of high specificity NTRIs is not surprising, in that initiations are almost always followed by repairs. In this case the predictor variable repair is a reaction to the initiation; it occurs afterwards. Giver P2 OI OR is negatively correlated; when givers repair rather than initiate in position two this leads to fewer high specificity follower NTRIs. This suggests an antagonistic relationship between the less common other-repair and other-initiation. It has been argued that there is a preference for initiation over repair (Schegloff et al., 1977); here it seems that correcting the other speaker reduces the negotiation of meaning that initiating repair creates. Giver sentential ellipsis increases with high specificity follower NTRIs; this may be due to position three OISR contributions being within this category. Giver ‘none’ or the absence of elliptical phenomena is negatively correlated; the use of elliptical contributions by the giver is positively related to the use of high specificity NTRIs by the follower. This suggests that the use and understanding of elliptical contributions requires checking and negotiating with any other participants through highly specific NTRI clarification requests. Giver vague anaphors are also negatively correlated; they of all anaphor types are associated with

fewer high specificity NTRI clarification requests.

Table 30: Stepwise multiple regression of predictors of ‘presumed shared context’ using ‘repair’ categories (only significant predictors are included)

variable	Multiple <i>R</i>	<i>B</i>	Standard error <i>b</i>	Beta	<i>t</i>	Significance. of <i>t</i>
All P1 Art.	0.35	4323.60	1021.72	0.35	4.23	.001

All P1 Articulation repairs was entered first and explained 12.4% of the variance in all ellipsis occurrences ($F_{1,126} = 17.91$, $P < .001$). No other significant predictors were found.

The explanation behind this relationship is not clear; it may be that the use of anaphora and other elliptical contributions cause a ‘stumbling’ over the current contribution, but this would need further investigation.

Table 31: Stepwise multiple regression of predictors of ‘repair’ categories’ from ‘presumed shared reference’ categories (only significant predictors are included)

variable	Multiple <i>R</i>	<i>B</i>	Standard error <i>b</i>	Beta	<i>t</i>	Significance. of <i>t</i>
All Ack.	0.49	-1733.15	253.52	-0.50	-6.84	.001
All Cataphora	0.55	16400.73	4846.77	0.25	3.38	.001
All Rejection	0.58	11216.01	4519.91	0.18	2.48	.014

All Acknowledgements were entered first and accounted for 24.2% of the variance ($F_{1,126} = 40.17$, $P < .001$). Then respectively all Cataphora (6.2%; $F_{1,125} = 11.08$, $P = .001$) and all Rejections (3.3%; $F_{1,124} = 6.16$, $P = .014$).

Repair is firstly negatively associated with acknowledgements; this is presumably due to a ‘right,ok’ type contribution signalling that there has been no need for the ‘trouble source – initiation – repair’ sequence^{22 23}. Cataphora may be a predictor of

²² This may entirely take place in position one or the longer P1-P2-P3 sequence.

²³ Although the sequence is quite often finalised by an acknowledgement which acts as a signal to continue.

repair due to its nature; the use of the pronoun before the descriptive noun may lead to confusion in both expressing and understanding a contribution. Rejection NSUs (especially in the context of the map task) indicate that there is a problem with a contribution; this is when repair sequences are introduced to fix the problem in understanding.

Table 32: Stepwise multiple regression of predictors of ‘anaphora’ from ‘repair’ categories (only significant predictors are included)

variable	Multiple <i>R</i>	<i>B</i>	Standard error <i>b</i>	Beta	<i>t</i>	Significance of <i>t</i>
F P1 Art.	0.26	0.33	0.11	0.25	2.88	.005
F P1 Form.	0.33	0.25	0.11	0.20	2.32	.022

Follower position one Articulation was entered first and accounted for 6.9% of the variance ($F_{1, 126} = 9.38, P = .003$). Follower position one Formulation accounted for a further 3.8% ($F_{1, 125} = 5.37, P = .022$).

That both forms of follower position one repair are the predictors of anaphor use is suggestive, but curious. From Chapter 4 it can be seen that givers use more anaphora than followers (due to the nature of the task) and it may be that givers are generally creating and using anaphoric references which followers have some degree of trouble accessing. However, these self-repairs in position one do not indicate a trouble in understanding of the shared context (that would be demonstrated through NTRIs); rather, they seem to make creating and expressing a contribution more difficult.

6.2.2 NTRIs and NSU questions

Are these measuring the same thing or not? Are different NTRI specificities capturing different ellipsis categories? Evidence seems to be inconclusive, shown in Table 33 below (all variables adjusted for words).

Table 33: Significant correlations between ‘Giver NTRI’ specificity levels 1-5 and ‘presumed shared context’ categories

NTRI type	Ellipsis category	Pearson r	p
GNTRI spec1	G slu	.245	.005
GNTRI spec1	Gnsuq	.201	.023
GNTRI spec2	Gslu	.207	.019
GNTRI spec3	Gslu	.424	.000
GNTRI spec3	Gnsuq	.209	.018
GNTRI spec4	Gslu	.244	.006
GNTRI spec4	Gce	.403	.000
GNTRI spec5	Gslu	.288	.001
GNTRI spec5	gnsuq	.222	.012

Table 33 shows that sluices are most commonly associated with giver NTRIs; this is intuitively clear due to requests for clarification frequently containing wh- words. However, giver NTRIs were relatively rare compared to follower NTRIs and are only given above for completeness. Followers contributed more NTRIs due to the task; these higher frequencies may be more representative of the relationship (only sig results given):

Table 34: Significant correlations between ‘Follower NTRI’ specificity levels 1-5 and ‘presumed shared context’ categories

NTRI type	Ellipsis category	Pearson r	p
FNTRI spec2	fchk	.244	.005
FNTRI spec4	fce	.239	.006

Here it seems that ellipsis categories and NTRIs coincide; ‘check’s can be comparable to level 2 NTRIs and clarification ellipsis contains a repeat of the trouble source as in specificity level 4. It may be surprising that sluices are not correlated; this may be due to clarification requests being phrased as complete sentences, rather than the fragment that the category ‘sluice’ requires. Clarification requests being

given as sentences could explain the overall low association between NTRIs and NSU questions.

6.2.3 Summary of results

From the regression analyses given above we can see that various measures of communicative process and content predict accuracy in the map task to different extents. In the first analysis using unadjusted totals two predictor variables emerged; 'follower NSU polar answers' and 'follower NSU rejection' each explaining 12% of the variance. For the analysis using standardised data adjusted for word totals, 'Giver instruct' from the DGA analysis emerged as the main predictor variable. A positive relationship was found between all ellipsis occurrences and position one articulation repairs. The major predictor of all repair occurrences was NSU acknowledgements, accounting for 24% of the variance. The most specific NTRI type was found to have main predictor variables of P3 OI SR, P2 OI OR (negatively correlated), Giver sentential ellipsis, Giver 'no ellipsis' and Giver 'vague anaphor' (the last two negatively correlated). The relationship between NTRIs and NSU questions such as sluices, clarification ellipsis etc is not strong although 'clarification ellipsis' is associated with NTRI specificity level 4. The speed at which the participants talked was examined and was found to be associated with categories from both the repair and ellipsis codings; speed of talk was negatively correlated with accuracy.

6.3 Preliminary conclusions

Accuracy in the map task can be understood as a function of successful communication, negotiated through processes that can be identified through repair and ellipsis coding categories. It appears that the availability of eye contact does not impact upon the use of the repair sequence or the mechanisms of grounding to a great extent in spite of theories emphasizing the importance of gaze and visual cues. Follower P3 SI SR appears more frequently in the audio-only condition and is also a (relatively weak) predictor of accuracy.

6.3.1 Review of correlational results

The analyses presented and discussed here have shown that the original DGA coding as well as the repair and ellipsis coding all capture a variety of dialogue processes. There is some crossover between DGA and repair/ellipsis categories. There is evidence for the dialogue processes involved with task accuracy, or success. Correlation analyses identify those variables that are related in some way to accuracy scores. Regression analyses identify predictor variables of the accuracy scores. These two statistical approaches are equally valid in understanding the underlying processes that are contained in the dialogues. It is possible to identify variables that appear to affect communicative success. Some of these are non-intuitive, such as more instructions leading to worse performance. Others are more easily understood, such as more rejections leading to worse performance.

6.4 Factors affecting successful interaction

From the results that have been presented and the speculations upon the causal relationships responsible, it is possible to suggest different factors that impinge upon the success of the task (assuming all other factors are equal):

1) Quantity of instruction

More instructions reduce the task success.

2) Complexity of question

Yes/no polar questions and answers reduce misunderstanding and improve task success, opposed to open-ended questions.

3) Strength of acknowledgement

Weak backchannels ('mmh') are less helpful than strong ('I understand').

4) Clarity of purpose

Difficulties in formulating a contribution contribute to low success.

5) Continued attention and monitoring

Monitoring such that mistakes or other trouble can be rectified improves success.

6) Acceptance of prior turn

Disagreements ('NSU rejections') demonstrate a lack of grounding and

reduce task success.

7) Completeness of information given

Unclear contributions that require further questions (not NTRI or clarification) indicate poor mutual-engagement or understanding and low task success.

Whether these factors or dimensions are applicable to other task-related or naturally occurring dialogue can only be speculated upon; in naturally occurring dialogue especially there are further variables that will affect communicative success.²⁴

6.4.1 Further dimensions

Some speculation can also be made about the factors of role, medium and familiarity. For tasks with a definite solution, if the task role is to instruct another person a noticeable pattern of repair and ellipsis use is found. This is also true for understanding and following another person's instructions. Medium and familiarity also demonstrate patterns of repair and ellipsis.

1. Instructing and explaining

Formulating and speaking instructions leads to a high level of self-monitoring and self-correction.

2. Understanding and following instructions

Following instructions increases the level of repair initiations and to a lesser extent other-repair.

3. Interacting with a familiar partner

A familiar dyad will use more repair initiations; they question each other more frequently.

4. Making eye contact

When eye contact is not possible there is an increase in the amount of clarifications, corrections and amendments made following a response.

²⁴ When dialogue is not goal-oriented 'success' may be moot.

6.5 Discussion

One of the aims of this research was to quantify the differences between contrasting forms of media. The repair/ellipsis coding did not discriminate between media through the occurrence of repair and ellipsis phenomena; the outcomes did not significantly vary between FTF and audio conditions. What might be possible reasons for negligible differences in task results between face-to-face and audio-only conditions? These reasons could be due to issues with the task, the coding protocols/dependent measures or other confounding factors.

Boyle et al. (1994) claim that for the HCRC maptask FTF is more efficient due to differences in number of turns, although there were no differences of accuracy. This is sufficient for Boyle et al. to claim:

“When subjects can see each other they can use visual signals, such as head nods and puzzled looks, as feedback, to provide evidence to their partner of their current state of understanding.” (Boyle et al., 1994; p. 15)

As this is presumably the case to an extent, it is curious that stronger repair and ellipsis differences were not found between media. Boyle et al. suggest that audio only dialogues had a range of compensatory behaviours, such as more interruptions and more backchannels than FTF dialogues; this allowed for equally accurate but longer dialogues. The efficacy of different media is harder to investigate due to the lack of medium difference regarding deviation scores. The efficacy of the two media is equal with regard to task accuracy; they both bring about the same result. If efficacy is to be measured by time taken, this is illustrated by the use of giver transition space repairs; follower position three self-initiated self-repairs; polar answers; acknowledgements; check questions. From this perspective it can be seen that the map task itself does not discriminate between eye contact or none.

From these data it cannot be said that audio-only is as efficacious as face-to-face communication; rather, that it makes some differences to the communicative process within the map task. If seen as a problem-solving task with a definite solution, this result replicates that of previous studies e.g. Chapanis (1986). It is arguable that the deviation scoring method was not precise enough; a different approach is described

in Davies (1998; see chapter 2). The ‘Incorrect Entity’ score examines the route negotiated around the various obstacles; rather than overall accuracy it examines how each individual landmark is dealt with. The score reflects two individual variables; whether the obstacle was shared by both giver and follower, and the size of error. The obstacles are the ‘pivotal’ elements to be examined, rather than the line between two obstacles being a few millimetres out (deviation was counted by laying a centimetre squared grid over the map and counting how many squares off the route the drawn line was). However, B. Davies (personal communication, April 2011) pointed out that there was a very high positive correlation between the Incorrect Entity and deviation scores; this suggests that similar results would be found even when landmark negotiation is concentrated on.

Two issues that could be examined in further research are regarding the navigation of turntaking, and more widely different media comparisons. Sacks et al. (1974) examined how turn-taking was regulated in naturally occurring conversation, when the underlying mechanisms are altered (e.g. the delay that occurs when making international phone calls) a very different distribution of repair and ellipsis would be predicted. More divergent media contrasts could be drawn between (e.g.) synchronous and asynchronous media, and multi-party dialogue FTF and mediated.

Although medium effects on interactive processes were not found in these studies there have been previous findings. Newlands et al. (2000) provide evidence that both the content and process of communication change due to the medium. Utilising the ‘Travel Game’ (Anderson et al., 1996), DGA coding showed that three games were more prevalent in a videophone condition than FTF; ‘explain’, ‘align’ and ‘query-YN’. The arbitrary nature of the DGA categories becomes an issue when trying to explain what these categories are identifying within communication; it is more useful to distinguish between evidence of understanding and artefacts of the specific task (e.g. the ‘explain’ game does not necessarily offer any insight into the communicative structure or success).

Another maptask study shows that significant differences between media can be

seen; Anderson et al. (2002) found that landmark referring terms were shortened in a VMC condition under time pressure, but not in FTF. Louwerson et al. (2007) provide evidence that the intentions of the speaker are correlated with eye gaze, facial expressions and pauses.

Further evidence comes from Anderson et al. (1999) examining 3-way map tasks (two givers, one follower). Anderson et al. contrasted the results of two and three person map tasks, also utilising media conditions (FTF versus videoconferencing). It was found that accuracy (deviation from the route) does not significantly vary over conditions, but for both media conditions the three party dialogues contained more turns and more words. In the three-party dialogues there were two information givers; the maps they each had slightly varied in the landmarks listed, although completely compatible with the others. To understand the difference in the three party dialogues, Anderson et al. examined the transcripts in order to determine exactly what was contained in the extra information exchange. It seems that the extra talk was used to allow all three participants to understand where the various missing landmarks were. This result was summed up by Carletta et al. (2002):

“The key lies in the establishment of common ground. (...) The landmarks that generated the extra discussion in the three-person groups were those known by one instruction giver and not the other. (...) These groups were establishing common ground for all the group members even when they did not need to. This suggests that in three-person groups, the norm is to aim for complete mutual-understanding, just as in dialogue.” (Carletta et al., 2002; p13).

Considering the difference in the repair profiles found in the BNC sample and the map task, it may be that the aim of task-oriented dialogue differs from unfocused naturally-occurring dialogue; it is questionable whether ‘the norm is to aim for complete mutual-understanding’. However, the importance of being able to negotiate shared understanding is clear. This can also be illustrated by the potential difficulties inherent in one-way communication, such as following a recipe in a cookbook or assembling flat-packed furniture. Negotiation is crucial to successful mutual-understanding of context.

The three-way map tasks would be a good starting point for an exploration of how anaphoric reference and repair is negotiated in multi-party dialogue. The patterns of frequencies found in the current studies illustrate how dyadic interaction is affected by role, medium and familiarity; this may change significantly in multi-party dialogue.

6.6 Review

Task oriented dialogue is different to naturally occurring dialogue in terms of the variables measured here. Giving instructions and exchanging questions and answers influence the process of interaction. Medium and familiarity are shown to have some affect on interaction, but the overall picture is that eye contact or mutual gaze is relatively unimportant in comparison to an audio channel. Familiarity does have some impact, such as frequency of questioning or challenging your interlocutor, but this makes little difference to outcome. The categories of repair and ellipsis outlined in the previous chapters allow a more complex understanding of the map task dialogues than has previously been available through dialogue games analysis. Critical variables such as formulation self-repairs in position one have not been previously identifiable through DGA and 'disfluency' analyses.

Chapter 7

Conclusions

7.0 Recap of thesis

This thesis presented evidence from research investigating the process of communication in dialogue and how this related to mutual understanding. In this chapter the basic position and findings of the thesis will be recapitulated. These findings are then discussed in the broader context of theories of communication. Further potential applications of this methodology are discussed, especially in terms of ‘Theory of Mind’ (Premack and Woodruff, 1978); how this relates to repair in dialogue and may be useful in examining communication in clinical populations.

The research presented here investigated aspects of dialogue which index mutual-understanding, the process by which speakers ensure that they share the content and meaning of the salient topics and references. It was argued that mutual-understanding of shared references and information is achieved through the process of repair and demonstrated through the elliptical use of elided material and pronominal reference and that these processes can be identified and measured in a way that is valid and reliable.

The analyses in this thesis found that repairs and ‘elliptical’ contributions can both be reliably extracted from transcripts using the protocols presented here. Anaphora and NSUs can be used to index mutual accessibility of context. Repairs and ellipsis index patterns of understanding and demonstrate different distributions over task type and familiarity. Eye contact and gesture appear to affect the mechanisms underlying shared understanding in terms of the repair and ellipsis categories used; some are significantly more prevalent in audio-only interaction. Complex patterns of correlations appear between repair and ellipsis categories. Some of the phenomena investigated predict accuracy in the map task. These are explained in the factors or

dimensions put forward in the last chapter.

In Chapter 2 Schegloff's (1993) arguments against the quantification of dialogue phenomena were outlined. Schegloff aimed his arguments at three areas: instances of an occurrence (e.g. laughter), the domain the instances are taken from (e.g. per minute), and the context from which the quantity is taken (e.g. informal conversation). Although counting 'laughter per minute' is agreed to be invalid, repair initiation and its response was acknowledged to be an exception by Schegloff. Self-initiation of repair was argued to be potentially possible at any point in talk (as are NTRIs) and identifiable in some cases where the self-edit takes place during the presentation of a contribution. Schegloff argued that conversational phenomena are not relevant when taken out of context; this thesis has argued that instances of repair are indices of mutual-understanding and given the nature of the HCRC map task, occur within the same context and this validates their quantification. Was quantification inappropriate for the BNC analysis? Firstly, the BNC sample was a random collection of dialogue under many different circumstances, avoiding the problem of only including dialogue from one particular scenario. Secondly, without a contrasting sample of dialogue it would not be possible to evaluate the relative occurrence of repair in the map task dialogues. Overall, the differences in the repair profiles over the different conditions illustrate the way that repair is utilised in different circumstances, in a way that has never been investigated previously. Conversation Analysis is a useful methodology for examining the structures and constituents of dialogue, but the findings have not been applied in a systematic way to different forms of dialogue. Although practitioners of CA may disagree with the (mis)use of the CA repair studies, quantification of any phenomena is fundamental to the systematic comparison of conditions.

7.1 Statistical results

In chapters 4, 5 & 6 various statistical results have been presented from the corpus analysis, the most relevant of which are summarised here. The map task corpus shows a very different repair profile to a sample from the BNC. 'Givers' in the map

task used more position one and three repairs; ‘followers’ used more position two initiations and repairs. With regard to medium, Follower P3 SI SR was significantly higher in the audio-only condition; Giver P1 TS was higher in the eye-contact condition. Follower P3 SI SR was found to positively correlate with P1 Articulation and P1 Formulation. There was a clear tendency to use high specificity NTRIs; NTRIs were in almost every case followed by a P3 OI SR from the interlocutor.

The indices of presumed shared context also showed a different profile to that of a sample from the BNC; for example there were fewer uses of exaphora in the map task, but more polar answers and acknowledgements. The two task roles also demonstrated different elliptical contribution use; givers used more anaphora while followers used a large number of non-sentential utterances (NSUs)²⁵. Polar answers, check questions and acknowledgements occurred more often in the audio-only medium condition. Follower high-specificity NTRIs were found to correlate positively with endophora, cataphora, polar answers and acknowledgements; negatively with giver ‘none’ (non-elliptical contributions). Polar answers negatively correlated with deviation from the route, and positively with acknowledgements.

Familiarity was one of the experimental factors in the map task design; some differences were found. In the repair analysis it was found that familiar pairs used significantly more P2 NTRIs. For the ‘ellipsis’ analysis it was found that familiar pairs used more endophora, polar and prompted answers, sluices, checks, non-sentential questions and rejections.

It may be useful to examine what has been measured by these protocols. Is it that repair measures ‘mutual-intelligibility’ and ellipsis ‘context’? ‘Intelligibility’ is a slightly fuzzy concept (e.g. deafness vs. using an obscure word), as is ‘mutual-engagement’, although the latter may be more appropriate for describing repair or initiations in positions two or three (to a lesser extent for position one, as the repair may be either due to self-monitoring or signals such as puzzled looks from an

²⁵ 75% of all contributions

interlocutor²⁶). In the map task corpus patterns of repair illustrate most clearly the intentionality of the interlocutors; in the BNC instances of NTRIs (for example) seem to be occurring arbitrarily (or at least unpredictably) rather than every debated instruction being cleared up to the satisfaction of both parties. Interestingly, in the HCRC map task corpus although the giver appears to use referring expressions such as anaphora, these are not reciprocated; followers tend to use specific terms, not abstracted referring terms. If this generalises to different aspects of communication, it may have implications for models of interaction. Established referring terms are thought to be preferred; when a term (e.g. anaphoric reference) is used, a listener expects this term to be consistent, and altering the referring term causes problems (e.g. Shintel and Keysar; 2009).

7.2 Models of interaction

In Chapter 2 some approaches to interaction were outlined, both experimental (e.g. Dialogue Games Analysis) and theoretical (e.g. Collaborative theory). These theories and approaches were described in order to illustrate previous attempts to understand or analyze dialogue; the analysis of mutual understanding presented in this thesis was not intended to operationalise and test these approaches. However, it is possible to attempt to interpret the mutual understanding data from the HCRC map task coding in terms of these approaches, with the following caveats.

The repair/anaphora and ellipsis coding were meant to contrast with previous approaches, not test them. The purpose of introducing them in Chapter 2 was to illustrate how they fail to present a general, testable account of the processes within interaction that create understanding. Testing the previous approaches would involve operationalising the concepts or examining the transcripts in particular aspects e.g. interruptions; this was not done as it was not the purpose of the thesis. Evaluating previous approaches through the data given by the repair/anaphora and ellipsis coding may lack validity; only very tentative conclusions could be made by evaluating e.g. grounding through frequency of non-sentential acknowledgements.

²⁶ This difference confounds to some extent medium effects that might be expected.

Two potentially opposing strands of thought emerged from previous studies:

- i. Audio is critical to task success, video and audio combined is no better (e.g. Chapanis' studies). This is dependent upon the task; although two task roles were analyzed separately the success of the task was a joint product – success cannot be attributed to either the giver or follower individually. No medium difference would be predicted.
- ii. Social cue theories argue that visual information supports the interaction process through organising turn taking and content through attention and head nods. Also task dependent, a medium difference would be predicted, all other things being equal. The social cue/non-verbal communication theories developed into Media Richness Theory and for a task such as the map task would predict that FTF would be more successful than audio.

It may be that only affective information is the critical variable in studies of medium differences; this is down to the type of task used and would not have been an advantage in the map task. It is possible that turntaking was more organised in FTF but this data was not available in the transcripts used for coding.

One influential theory described in Chapter 2 was Clark's Collaborative/Grounding theory. This was previously touched upon in light of the dialogue coding data in Sections 6.1.5 and 6.1.6. The grounding or collaborative model stated that acknowledgement is one of the types of positive evidence of understanding, argued to be preferred over negative evidence such as requests for clarification. This is the only one of the four types of positive evidence put forward by Clark and Brennan that is suitable to be examined from the data used in this study; however, in Clark and Brennan's view this is the second strongest evidence of understanding and so may be thought of as a potentially strong indicator of understanding or 'grounding' in the collaborative theory. With the data presented in this thesis it could be argued from the collaborative theory that the category of 'NSU acknowledgements' would be positively correlated with accuracy; any type of repair initiation or associated types of NSU questions would be less preferred although still associated with mutual

understanding and increased performance in tasks. Two main factors were argued to affect grounding; purpose (here, task role) and medium with inherent costs. Accuracy in this task does not differentiate between the two task roles as it is a joint process. Costs would vary over the two medium conditions; it was found that NSU acknowledgements did not change. Correlational results show that NSU acknowledgements are actually significantly associated with a less successful task outcome; they are also negatively correlated with 'all repairs'.

The principle behind Media Richness Theory is that as modes of simultaneous expression increase, for a given task communication and communicative/task success increases due to the richness of the medium. Due to the task not varying over medium conditions it would be predicted that FTF would have an advantage and lead to more accurate routes in the map task. Is there evidence that FTF has an advantage over audio-only due to richness/more social cues? The differences outlined in the previous chapters show that there are few significant differences; ('vague anaphor' higher in FTF condition; NSU polar answer, Follower P3 SI SR both higher in audio condition). These results do not suggest an advantage; rather a difference in the communicative process and this would appear to contradict media richness theory's basis (although media richness theory should be evaluated with respect to different task types and types of outcome e.g. the contrasting outcomes from creative brainstorming, negotiation or logical problem solving).

One development of 'richness' theory, Media Synchronicity Theory may make more specific predictions. Familiarity and associated group norms are thought to be advantageous and can enable tasks to be carried out equally well with less 'rich' information. Also, levels of 'conveyance' and 'convergence' depend upon task and familiarity; an unfamiliar pair facing a new task will be at a disadvantage. Conveyance is argued to be built through questions, proposals and contextualisation; of the available data, this would presumably be measurable from the NSU question categories and possibly NSU statements (rejection; modifier; continuation). Convergence comes about through mutual agreement; it is not clear if any of the coding categories measure this directly although it may possibly be indexed by a

negative correlation of task success to the 'NSU rejection' category. From this it would be predicted that familiar pairs in both medium conditions would be more successful in the task; familiar pairs would demonstrate higher levels of NSU questions, modifiers, continuations and lower NSU rejections. With regard to accuracy in the map task, familiarity did not significantly differ from the unfamiliar condition (Boyle et al., 1994). NSU rejections did not significantly differ over familiarity conditions but there was a significant increase in familiar Giver's NSU Sluice and overall NSU questions. Overall P2 NTRI repair types approached significance (more frequent in the familiar condition) and further examination showed these to be the two highest specificity Follower NTRI types. Familiarity did not affect the frequency of any of the NSU statement types. Normative behaviour associated with familiarity had some effect on the use of NSUs but not in ways predicted by MST.

The coding categories that make up Dialogue Games Analysis were originally created in order to explicitly label the processes occurring during task oriented dialogue. The DGA categories that explicitly supported mutual understanding in the HCRC map task corpus were examined by Doherty-Sneddon et al. (1997; Study 1). Doherty-Sneddon et al. examined medium differences in 'initiating moves' and concluded that the audio-only condition required more 'check' and 'align' moves in order to compensate for the lack of visual feedback. Is this supported by the dialogue coding presented in this thesis? Regarding feedback and feedback requests (i.e. 'check' and 'align' type actions) there is no evidence for differences across the medium differences; these would have been identified by follower NSU check/follower clarification ellipsis/follower NTRI categories for the DGA 'check' category, and giver NSU check/giver clarification ellipsis for the DGA 'align' category. These did not significantly differ over the medium conditions.

Boyle et al. (1994) argue that the collaborative theory and the CA model of dialogue have opposing views with regard to how successful dialogue can be judged. Boyle et al. claim the CA model contains an implicit notion of task success based upon smooth changes in turn-taking with few gaps or overlaps; the goal of dialogue is to

regulate turn-taking:

“When a conversation breaks down the problem can be attributed to a failure in the turn taking procedure (Cutler and Pearson, 1986)” (Boyle et al., 1994; p 3)

This idea is contrasted with the collaborative model, which sees mutual-understanding being achieved through grounding. Dialogue success is based upon collaboration that facilitates grounding; turn-taking emerges from this and interruptions to the turn-taking process do not necessarily interfere with grounding (Boyle et al., 1994). The role of eye contact is not explicitly explained in these accounts; from Clark and Brennan (1991) it can be inferred that the constraints of copresence and visibility would lead to common ground being negotiated verbally to a greater extent where eye contact is not possible.

For the map task, in terms of Clark and Brennan’s constraints for grounding the eye contact condition has (limited) copresence and visibility over the audio-only condition. This would suggest that in the eye contact condition there would be fewer instances of the giver verbally checking the followers’ understanding. This would be demonstrated through NTRIs, sluices, clarification ellipsis, checks and NSU questions; also check and align from the DGA coding. There is some support for this; there were more NSU checks in the audio-only condition. From the repair protocol the results were slightly different; rather than fewer checking questions in the eye contact condition there were more instances of transition space and position three self-initiated self-repairs. These types of repairs can depend to some extent on observing expressions of puzzlement or confusion on the other participant’s face.

7.3 Implications for models of interaction

It is possible to apply these findings to a more global conception of communication. Mills (2007) describes the contrast between the collaborative model and interactive alignment model, concluding that a model is needed which ‘emphasizes the role of interactive repair as a key process underlying the development of semantic co-ordination’ (Mills, 2007; p. 5). Mills goes on to state that “it is the role of negative feedback, in particular participants’ ability to question and clarify each other’s

location descriptions that is fundamental to the development of more abstract descriptions.” This conception of negotiation of shared understanding as a ‘negative’ does not seem correct; although here Mills is actually arguing that it is a ‘positive’ or facilitating factor in the formation of abstract references. The elliptical NSU category ‘rejection’ could be thought of as an unhelpful ‘negative’, but Fernandez (2006) points out this is most often given as a ‘help rejection’ whereby a contradictory response is followed by a (unprompted) further answer or information²⁷. It appears to be very rare for dialogue to be thwarted by ‘negative’ contributions; presumably these would occur in situations such as a policeman questioning a reticent and laconic criminal. In these cases of being deliberately unhelpful it may be safe to assume that Grice’s maxims and Clark’s collaborative model are not functioning.

Every single map task transcript from the HCRC corpus contained substantial amounts of repair and ellipsis. The approach taken in Clark et al.’s ‘collaborative theory’ is that as mutual-understanding increases, dialogue contributions become shorter as referring terms become part of the common ground. Clark and Krych (2004) note that various elliptical phrases can be used to establish common ground, from continuers (‘uh-huh’, ‘yeah’) or assessments (‘gosh’) to establishing shared attention through deictic expressions such as ‘this’, ‘that’, ‘here’ and ‘there’. Healey et al. (2008) demonstrated the basic concept and viability of quantifying elliptical phenomena as a quantitative index of mutual-accessibility of context. They showed that the frequency of use of cross-speaker elliptical expressions in online chat varies systematically depending on whether communication is ‘local’ i.e. within a single chatroom or ‘remote’. However, the coding of presumed shared context/‘ellipsis’ in the Healey et al. study did not follow an explicit protocol. It relied mainly on the distinctions made by Fernandez et al. (2004) but specific measures of reliability and validity were not calculated.

²⁷ Although NSU rejections do negatively correlate with accuracy in the multiple regression, discussed in the last chapter.

Healey (2004) suggests that, “Explicit negotiation and repair are relatively rare and do not provide a general account of co-ordination” (p. 2). This may be the case for a ‘general’ model of naturally occurring dialogue, but not necessarily with regard to task-oriented dialogue or other communication with specific problems, such as a non-native speaker of a language interacting with a native speaker. Repairs in this case are so common that they “constitute the basic pattern of conversation between the speakers” (Ulichny, 1997; p. 233); they cannot be described as ‘relatively rare’.

It appears that the nature of the dialogue/task is crucial to the degree to which mutual-understanding must be explicitly negotiated. As far as the results presented in this thesis can be generalised, it appears that when precise information needs to be imparted to an interlocutor, explicit negotiation of understanding becomes a crucial part of dialogue. To some extent the medium (eye contact or none) indicates the type of repair that predominantly emerges (more P1 TS with eye contact; more P3 SI SR with no eye contact). It also appears that the relationship between medium and communicative process is affected by the purpose behind the communication.

7.4 Further applications of the methodology

The methodology presented in this thesis allows for the identification of dialogue phenomena which have been argued to index mutual understanding. This could be of use in studying clinical populations such as people diagnosed with schizophrenia or autistic spectrum disorders. Firstly it may be useful to examine if measureable aspects of mutual understanding occur in significantly different frequencies within separate clinical populations; secondly it may be that aspects of mutual understanding index to some extent ‘Theory of Mind’ (ToM) (Premack and Woodruff, 1978). ToM refers to the ability to attribute mental states and knowledge to others; it is also referred to as metarepresentation. The lack of ToM has been argued to be a feature of people diagnosed with autistic spectrum disorders (Baron-Cohen et al.; 1985) and schizophrenia (Frith, 1992). If ToM is thought to be indexed through the use and understanding of both elliptical contributions and repair this would enable dialogue transcripts from various clinical populations to be contrasted

in terms of evidence for ToM.

The version of the repair protocol published in Healey et al. (2005) was used by Themistocleous et al (2009) in a study examining conversational repair in psychiatric consultations; it was found that patients used more self-repair (presumably to clarify themselves and avoid being misunderstood) and psychiatrists initiated repair more in order to reduce misunderstandings.

“In this study, rather than trying to define ‘good communication’ in some abstract sense, we focus on the details of how participants work to sustain the mutual-intelligibility of an interaction using the concept of repair. It is in this sense that Schegloff (1992) describes repair as the primary site of intersubjectivity in conversation. Note that, all things being equal, the implication is that higher levels of repair are indicative of people working harder to establish mutual understanding.”
(Themistocleous et al., 2009; p167)

Themistocleous et al. appear here to present ‘mutual-intelligibility’ as synonymous with ‘mutual-understanding’; that has not been the case in this thesis as mutual-intelligibility may be understood to be referencing factors such as a thick accent or bad handwriting. Despite this potential problem with terminology the approach is similar to that taken in this thesis; rather than looking for communicative ‘success’ or other debatable aspects of interaction the contents of the interaction are examined for evidence of mutual understanding.

“The aim of this study is to overcome the limitations of assessing doctors’ styles of interacting by looking at what patients and doctors do with each other’s talk. It exploits the structure of repair to gain an understanding of how patients and psychiatrists attempt to understand and make their talk understandable to each other.”
(Themistocleous et al, 2009; p168)

McCabe et al. (2005) claim that examining use of third position repair refutes the idea that ToM is impaired in people diagnosed with schizophrenia. McCabe et al claim that there is a direct connection between metarepresentation/ToM and repair in position three:

“Instances of third position repair in our data would be strong evidence

that a patient is not relying on routinised interactional moves but can understand that their interlocutor has misunderstood a particular utterance and respond instantaneously to correct that misunderstanding. Moreover, it shows an ability to diagnose the nature of the misunderstanding.”
(McCabe et al., 2005; p. 5)

These two studies suggest that a quantitative study of dialogue taken from clinical populations may be of use in understanding the strengths and weaknesses of that population in terms of mutual understanding. Experimental lab-based studies will always suffer from a lack of ecological validity; direct examination of dialogue – especially naturally occurring – avoids the usual potential confounds. It is possible that such an approach would provide useful evidence in cases such as the current debate over whether children diagnosed with autistic spectrum disorders have an impaired ToM with associated problems in dialogue;

“...children with autism, a condition which has been claimed to be associated with an impaired ToM (Baron-Cohen et al., 1985) appear to experience communication failures more frequently than their typically developing peers (see Keen 2003). However, Volden’s (2004) study of the problem yielded ambiguous results... children with autistic spectrum disorder (ASD) performed similarly to controls in responding to requests for clarification.”
(Bosco and Tirassa, 2010; p.65)

A quantitative approach, based upon reliable indices of mutual understanding could provide relative frequency distributions of the various categories of understanding and trouble. This could provide data that has so far not been available and would demonstrate any difficulties found in the construction and negotiation of mutual understanding. Attempts to examine clinical populations such as children diagnosed with schizophrenia have been reliant on descriptions and non-standardised definitions of dialogue phenomena such as repair, for example Bosco and Tirassa (2010); Bosco et al. (2006);

“The symptomatology of schizophrenia... has been explained on the basis of a ToM impairment (Frith 1992). Children suffering from schizophrenia have been described as using self-initiated repair strategies like repetition, revision and fillers less frequently than normally

developed children (Caplan et al. 1996). Adults suffering from the same disorder have been described as attempting to self-repair their messages inadequately during a referential communication task (Leudar et al. 1992). (Bosco and Tirassa, 2010; p. 64)

Note that the identified differences are ‘descriptions’; there have not been quantitative studies using reliable coding protocols to examine these dialogue phenomena in contrasting populations. Bosco et al. (2006) also take a descriptive approach, and tie this into a ToM-based theory of how children use repair in communication. It is argued that in order to realise that communication is not being successfully achieved the knowledge states of others must be accessible; this is demonstrated through the use of repair and repair initiations. From this view the lack of repair and repair initiations indicates a ToM deficit.

“Feldman and Kalmar (1996) suggest that when an actor repairs a communicative failure, she usually adapts her repair strategy in order to take her interlocutor’s perspective into account. This means that the actor repairing a failure tries to imagine why her partner failed to understand her communicative intention or accept her request. The cognitive component underlying such an ability is, in Feldman and Kalmar’s view, the theory of mind. ...In line with Feldman and Kalmar’s findings, we argue that theory of mind plays a role not only in the ability to repair a communicative failure, but also in the ability to recognize it. In order to recognize a failure, the actor has to realize that she failed to modify her partner’s mental state in the desired way. In our view, the theory of mind underlies such an ability.”
(Bosco et al.; 2006)

7.5 Conclusions and further work

Measures of repair and ellipsis seem to be reliably measured by the protocols presented here. These measures appear to discriminate between various factors. These include differences in distribution in both the BNC and HCRC map task corpora, and various experimental measures within the map task. Some predictions of task outcome may be possible through measurement of repair/ellipsis. Some categories from the repair protocol, ellipsis protocol and DGA approach appear to measure the same things. The advantage of the repair/ellipsis approach is that there is no regard taken to either the content of the communication or the medium; it is

totally neutral in that respect and can be applied to any communicative act.

A potential drawback of the repair protocol may be that it does not have the flexibility of qualitative approaches such as CA, which can provide new developments based upon prior findings. For example, Schegloff et al. (1977) do not give definitions of the various repair types; exemplars are provided and their relevance is discussed. The protocol would require redesigning and retesting for reliability in order to encompass further phenomena. The approach outlined here also assumes that interaction can be captured by a transcript which can then have the protocol applied.

DGA categories have been used in studies of CMC; e.g. examining frequency of 'check' or 'align' categories in order to test the accumulation of common ground (for example Monk and Gale 2002; Convertino et al. 2008). It may be that application of the two protocols presented here would allow more precise analysis of the nature of any negotiation of common ground. A 'check' move from DGA could refer to a very precise question such as 'Do you understand that...' or a vague 'ok?'; 'align' moves similarly can be precise rephrasings (such as repeating a telephone number to ensure it is written down correctly) to candidate understandings ('Do you mean...') or less specific NTRI repeats of a single word or phrase. Monk and Gale (2002) show that both check and align categories vary over medium conditions and argue that this is due to 'full gaze awareness' (seeing where another person is looking); in this condition there are fewest checks and aligns. Monk and Gale suggest that grounding requires monitoring of the other and when possible this will be done through watching gaze and expressions rather than through speech. These results may be more useful for CMC design if it was known what types of queries were and were not needed in the various conditions; repair structure, anaphoric reference and use of ellipsis give a detailed picture of how dialogue has been used.

From these data it is clear that human-human interaction is highly adaptable; although the process of communication illustrated by the protocol codings was different across the conditions ultimately the task success did not differ. The question

then becomes one of gauging which conditions required less effort, in order to prescribe for future CMC design. Are there other channels or constraints that would be useful, based upon these data? It would seem that due to the lack of task outcome differences there would be no obvious additional non-verbal channels to add to ease the fixing of interactional trouble. When eye contact is not possible there is no access to continuers or trouble signals (nodding head or frowning) and this cannot be compensated for in a simple way. Other studies seem to suggest that more efficient collaboration can be made through knowing where another person is looking (Monk and Gale, 2002) or through shared control of a cursor/pointer (Doerry, 1995; Whittaker, 2002).

It may be that understanding of these phenomena is useful for the development of computer based dialogue systems. Not only do systems with appropriate self-repairing ability seem more natural (Skantze and Hjalmarsson, 2010; Aist et al., 2007), but a system that can revise its plan and make this obvious to a listener has an advantage in terms of communicative clarity, as the presumed preference for example (1.2) below demonstrates:

(1.1) System: I have two seats available. I have one seat available.

(1.2) System: I have two seats...uh no... one seat available.

(taken from Guhe and Schilder, 2002).

7.6 Final thoughts

Both repair phenomena and elliptical contributions are ubiquitous in dialogue. The approach described in this thesis allows for complex relationships between different aspects of mutual-understanding to be explored. When the HCRC map task corpus is examined for use of these phenomena, patterns emerge that illustrate some of the factors associated with those phenomena. The methodology presented here can be applied to any communicative act regardless of medium, and could be used to investigate the patterns of repair and ellipsis use in larger groups, including the behaviour of bystanders and overhearers. Further potential uses could include the dialogue phenomena found in clinical populations.

References

Aist, G, Allen, E. Campana, C.A. Gomez Gallo, S. Stoness, M. Swift, and Tanenhaus, M. K. (2007). Incremental dialogue system faster than and preferred to its nonincremental counterpart. In *Proceedings of the 29th Annual Conference of the Cognitive Science Society*, 2007.

Alcántara, M. and Bertomeu, N. (2005). Ellipsis in Spontaneous Spoken Language. *17th European Summer School on Logic, Language and Information (ESSLLI 2005)*, August 8-12, 2005, Edinburgh, Scotland.

Alexandersson, J., B. Buschbeck-Wolf, T. Fujinami, M. Kipp, S. Koch, E. Maier, N. Reithinger, B. Schmitz & M. Siegel (1998) Dialogue acts in VERBMOBIL-2 (second edition). *Verbmobil Report 226*. DFKI, Saarbrücken.

Alexanderson, J., Maier, E., and Reithinger, N. (1995). A robust and efficient three-layered dialogue component for a speech-to-speech translation system. In *Proceedings of the seventh European meeting of the association for computational linguistics*, pp188-193.

Allen, J. & Core, M. (1997). *DAMSL: Dialogue Act Markup in Several Layers (Draft 2.1)*. Technical Report, Multiparty Discourse Group, Discourse Resource Initiative, September/October 1997.

Allen, J., Schubert, L., Ferguson, G., Heeman, P., Hwang, C. H., Kato, T., Light, M., Martin, N., Miller, B., Poesio, M. and Traum, D. (1994). The TRAINS project: A case study in defining a conversational planning agent. *Technical Report 532*, Department of Computer Science, University of Rochester, New York.

Ames, M.G., Go, J., Kaye, J., and Spasojevic, M (2010). Making love in the network closet: the benefits and work of family videochat. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. ACM, p. 145–154.

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., Isard, S., Kowtko, J., McAllister, J. Miller, J., Sotillo, C., Thompson, H. S. & Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech*, 34, pp. 351-366.

Anderson, A. And Howarth, B. (2002). Referential form and word duration in video-mediated and face-to-face dialogues. In Bos, Foster & Matheson (eds): *Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002)*, 4-6 September 2002, Edinburgh, UK, Pages 13-20.

Anderson, A. H., Mullin, J., Katsavras, E., Brundell, P., McEwan, R., Grattan, E., & O' Malley, C. (1999). *Multimediating multiparty interactions*. Proceedings of INTERACT 99. IFIP.

Anderson, A.H., Newlands, A., Mullin, J., Fleming, A.M., Doherty-Sneddon, G., and Van der Velden, J. (1996). Impact of video-mediated communication on simulated service encounters. *Interacting With Computers*, 8 (2):193-206.

Anderson, A. H., O'Malley, C., Doherty-Sneddon, G., Langton, S., Newlands, A., Mullin, J., Fleming, A., and Van Der Velden, J. (1997). The impact of VMC on collaborative problem solving. In K. Finn, A. Sellen and S. Wilbur (Eds.), *Video Mediated Communication*, pp. 133-156. Mahwah, NJ: Lawrence Erlbaum.

Argyle, M., Lalljee, M. and Cook, M. (1968). The effects of visibility of interaction in a dyad. *Human Relations*, 21, 3-17.

Argyle, M., Lefebvre, L. and Cook, M. (1974). The meaning of five patterns of gaze. *European Journal of Social Psychology*, 4, 125-136.

Atkinson, J. and Heritage, J. (1984). *Structures of Social Action: studies in conversation analysis*. Cambridge University Press.

Bard, E. G. and Lickley, R. J. (1997) On not Remembering Disfluencies. In *Proceedings of Eurospeech 97*, Rhodes.

Bard, E. G., Lickley, R. J. and Aylett, M. P. (2001): "Is disfluency just difficulty?", In *DISS'01*, 97-100.

Baron-Cohen, S., Leslie, A.M. and Frith, U. (1985) 'Does the autistic child have a theory of mind?', *Cognition*, 21: 37-46.

Bortfeld, H., Leon, S., Bloom, J., Schober, M., & Brennan, S. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, 44, 123-147.

Bosco, F.M., Bucciarelli, M. and Bara, B.G. (2006) 'Recognition and recovery of communicative failures: a developmental perspective', *Journal of Pragmatics*, 38: 1398-1429.

Bosco, F.M. and Tirassa, M. (2010). Communication failure. In: *The Pragmatics Encyclopedia*, ed. L. Cummings (pp. 63-65). London and New York: Routledge.

Boyle, E., Anderson, A. and Newlands, A. (1994). The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language and Speech*, 37, 1-20.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, 44, 274-296.

Brown G, Anderson AH, Yule G, Shillcock R. (1984). *Teaching Talk*. Cambridge University Press: Cambridge.

Bunt, H. (2006). Dimensions in Dialogue Act Annotation. In *Proceedings of the 5th International Conference on Language Resources and Systems (LREC 2006)*,

Genova, May 2006, pp. 919–924.

Bunt, H. (2009) Multifunctionality and multidimensional dialogue semantics. In *Proceedings of the Workshop on the Semantics on the Semantics and Pragmatics of Dialogue (DiaHolmia)*, Stockholm, July 2007, pp. 3–14.

Burnard, L. (2000). *The British National Corpus Users Reference Guide*. <http://www.natcorp.ox.ac.uk/docs/userManual/> Retrieved 21st May 2012

Caplan, R., Guthrie, D. and Komo, S. (1996) ‘Conversational repair in schizophrenic and normal children’, *Journal of the American Academy of Child and Adolescent Psychiatry*, 35: 950-58.

Carletta, J. (1996). Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22, 249-254.

Carletta, J. (2008) Reliability Measurement: There's no safe limit. *Computational Linguistics* 34(3): 319-326.

Carletta, J., Anderson, A. H., and Garrod, S. (2002). Seeing eye to eye: an account of grounding and understanding in work groups. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, 9(1), 1-20, March.

Carletta, J. C., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G., & Anderson, A. (1996). *HCRC Dialogue Structure Coding Manual (HCRC/TR-82)*, Human Communication Research Centre, University of Edinburgh.

Carletta, J., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G. and Anderson, A. (1997). The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23, pp. 13-31.

Chapanis, A. (1975). Interactive Human Communication. *Scientific American*, 232, 34-42.

Chapanis, A. (1986). Interactive Human Communication. In I. Grief (ed.) *Computer supported cooperative work: a book of readings*, pp. 127-140. San Mateo, CA: Morgan Kaufman Publishers Inc.

Chapanis, A., Ochsman, R., Parrish, R. and Weeks, G. (1972). Studies in interactive communication: I. The effects of four communication modes on the behaviour of teams during cooperative problem solving. *Human Factors*, 14, 487-509.

Chapanis, A., Parrish, R., Ochsman, R. and Weeks, G. (1977). Studies in interactive communication: II. The effects of four communication modes on the linguistic performance of teams during co-operative problem solving. *Human Factors*, 19(2), 101-126.

Clark, H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H. and Brennan, S. (1991). Grounding in Communication. In L. B. Resnick, J. Levine and S. D. Teasley (Eds.), *Perspectives on socially shared cognition*. Washington: APA Press.

Clark, H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50(1), 62-81.

Clark, H. and Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, 13, 259-294.

Clark, H. and Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1-39.

Clayman, S. E. and Heritage, J. (2002). Questioning presidents: Journalistic

deference and adversarialness in the press conferences of Eisenhower and Reagan. *Journal of Communication*, 52, 749-775.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37-46, 1960.

Cohen, K. (1982). Speaker interaction: video teleconferences versus face to face meetings. In *Proceedings of teleconferencing and electronic communications*, 189-199. Madison: University of Wisconsin Press.

Cohen, P. R. (1984). The pragmatics of referring and the modality of communication. *Computational linguistics*, 10(2), 97-146.

Collard, P. (2009). *Disfluency and listeners' attention: an investigation of the immediate and lasting effects of hesitations in speech*. Unpublished PhD thesis, University of Edinburgh.

Colman, M., Eshghi, A. and P. G. T. Healey (2008). 'Quantifying ellipsis in dialogue: an index of mutual-understanding.' In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, June 2008 Columbus, Ohio: Association for Computational Linguistics.

Colman, M. And Healey, P. G. T. (2011). 'The distribution of repair in dialogue'. In *Proceedings of the 33rd Annual Cognitive Science Conference*, July 2011, Boston, Massachusetts.

Condon, S. L., and Čech, C. G. (1996) Functional comparison of face-to-face and computer-mediated decision-making interactions. In Herring, S. (Ed.), *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives* (65-80). Philadelphia: John Benjamins.

Convertino, G., H.M. Mentis, M.B. Rosson, J.M. Carroll, A. Slavkovic, and C.H.

Ganoë (2008) "Articulating common ground in cooperative work: Content and process," in *Proceedings of 26th Annual ACM Conference on Human Factors in Computing Systems*, 2008, pp. 1637-1646.

Cutler, A and Pearson, M. (1986). On the analysis of prosodic turn taking cues. In C. Johns-Lewis (ed.) *Intonation in discourse*. pp 139-155. London: Croom Helm.

Daft, R. L. and Lengel, R. H. (1984). Information richness: a new approach to managerial behaviour and organizational design. *Research in Organizational Behaviour*, 6, 191-233.

Daft, R. L. and Lengel, R. H. (1986). A proposed integration among organizational information requirements, media richness and structural design. *Management Science*, 32, 554-571.

Davies, B.L. (1998). *An Empirical Examination of Cooperation, Effort and Risk in Task-Oriented Dialogue*. Unpublished PhD Thesis. University of Edinburgh.

Davies, B.L. (2006). Testing dialogue principles in task-oriented dialogues: An exploration of cooperation, collaboration, effort and risk. *Leeds Working Papers in Linguistics and Phonetics*, No.11.

Dennis, A. R., Fuller, R. M., & Valacich, J. S. (2008). Media, tasks, and communication processes: A theory of media synchronicity. *MIS Quarterly: Management Information Systems*, 32(3), 575-600.

Dennis, A. and Valacich, J. (1993). Computer brainstorm: more heads are better than one. *Journal of Applied Psychology*, 78, 531-537.

Dennis, A. R. and Valacich, J. S. (1999). Rethinking Media Richness: Towards a theory of media synchronicity. *Proceedings of the 32nd Hawaii International Conference on System Sciences*.

Doerry, E. (1995), "*An Empirical Comparison of Copresent and Technologically-Mediated Interaction based on Communicative Breakdown*", PhD Dissertation, Univeristy of Oregon, Eugene.

Doherty-Sneddon, G., Anderson, A. H., O'Malley, C., Langton, S., Garrod, S., & Bruce, V. (1997). Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied*, 3, 105-125.

Eshghi, A. (2009). *Uncommon ground: the distribution of dialogue contexts*. Unpublished PhD thesis, University of London.

Eshghi, A. & Healey, P.G.T. (2009). What is conversation? Distinguishing dialogue contexts. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Amsterdam. Holland.

Feldman, C.F. and Kalmar, D. (1996) 'You can't step in the same river twice: repair and repetition in dialogue', in C. Bazzanella (ed.) *Repetition in Dialogue*, Tübingen: Niemeyer.

Fernandez, R. (2006). *Non-sentential utterances in dialogue: classification, resolution and use*. Unpublished Ph.D. thesis, University of London.

Fernandez, R. Ginzburg, J. and Lappin, S. (2004). Classifying Ellipsis in Dialogue: A Machine Learning Approach. In *Proceedings of the 20th International Conference on Computational Linguistics, COLING 2004*, pages 240–246, Geneva, Switzerland, 2004.

Fernández, R., Schlangen, D., and Lucht, T. (2007). Push-to-talk ain't always bad! Comparing Different Interactivity Settings in Task-oriented Dialogue. In *Proceedings of the 11th International Workshop on the Semantics and Pragmatics of*

Dialogue (DECALOG), pp. 25-31, Trento, Italy.

Finholt, T., Sproull, L. and Kiesler, S. (1990). Communication and performance in ad hoc task groups. In J. Galegher, R. Kraut, and C. Egido (Eds.), *Intellectual Teamwork: social and technological foundations of cooperative work*, pp. 291-326. Hillsdale, NJ.: Erlbaum.

Frederking, R.E. (1993). *Understanding Dialogue Ellipsis*. Technical Report CMU-CMT-93-142, Carnegie Mellon University, 1993.

Frith, C.D. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Hove, UK and Hillsdale, NJ: Lawrence Erlbaum Associates.

Furnham, A. (1982). The message, the context, and the medium. *Language and Communication*, 2, 33-47.

Galegher, J. and Kraut, R. (1996). Computer mediated communication for intellectual teamwork. In R. Rada (Ed.), *Groupware and Authoring*. London: Academic Press.

Gersick, C. J. G. and Hackman, J. R. (1990). Habitual routines in task-performing groups. *Organizational Behaviour and Human Decision Processes*, 1, 47, 65-97.

Ginzburg, J. (2012). *The Interactive Stance*. Oxford University Press: Oxford.

Goffman, E. (1981). *Forms of Talk*. Philadelphia: University of Pennsylvania Press.

Goffman, E. (1976). Replies and Responses," *Language in Society* 5:3 (1976), pp 257-313

Gordon, S. and Gordon, J. (1996). *Information Systems: A management approach*. Fort Worth TX: The Dryden Press.

Grice, H.P. (1975). 'Logic and conversation' In Cole, P. & Morgan, J. (eds.) *Syntax and Semantics*, Volume 3. New York: Academic Press. pp. 41-58.

Guhe, M and Schilder, F (2002). Incremental generation of self-corrections using underspecification. *Language and Computers*, 45(1):118–132, 2002.

Healey, P. G. T. (1999). "Accounting for Collaboration: Estimating Effort, Transparency and Coherence" in *AAAI Fall Symposium, Psychological Models of Communication in Collaborative Systems* pp. 54--59, 1999.

Healey, P. G. T, Colman, M. and Thirlwell, M. (2005). "Analysing Multi-Modal Communication: Repair-Based Measures of Human Communicative Co-ordination" in *Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems*, Dordrecht: Kluwer Academic.

Healey, P.G.T., White, G., Eshghi, A., Reeves, A.J. and Light, A. (2008). Communication Spaces. *Computer Supported Cooperative Work; 17*, 169-193.

Healey, P. G. T. and Thirlwell, M. (2002). "Analysing Multi-Modal Communication: Repair-Based Measures of Communicative Co-ordination" in *Proceedings of the International CLASS Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems* pp. 83--92, June 28th- 2002.

Heritage, J. (2004). 'Conversation Analysis and Institutional Talk.' In Robert Sanders and Kristine Fitch (eds), *Handbook of Language and Social Interaction*. Mahwah NJ, Erlbaum: pp. 103-146.

Heritage, J., Robinson, J., Elliott, M., Beckett, M. and Wilkes, M. (2007). Reducing patients' unmet concerns in primary care: The difference one word can make. *Journal of General Internal Medicine* 22 (10): 1429–1433.

Hollingshead, A. and McGrath, J. E. (1993). The whole is less than the sum of its parts: a critical review of research on computer-assisted groups. In R. A. Guzzo and E. Salas (Eds.), *Team decision and team-performance in organisations*. San Francisco: Jossey-Bass.

Howell, D. C. (1994). *Statistical methods for psychology*. Duxbury Press.

ISO DIS 24617-2 (2010). *Language resource management – Semantic annotation framework (SemAF), Part 2: Dialogue acts*. ISO, Geneva, January 2010.

Kahneman, D. (1973). *Attention and effort*. New Jersey: Prentice Hall.

Keen, D. (2003) 'Communicative repair strategies and problem behaviours of children with autism', *International Journal of Disabilities, Development and Education*, 50: 53-63.

Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychologica*, 26, 1-47.

Kowtko J.C, Isard, S.D & Doherty, G. (1991). *Conversational Games in Dialogue*. Technical Report HCRC/RP-31. HCRC Publications University of Edinburgh.

Krauss, R. M. and Weinheimer, S. (1966). Concurrent feedback, confirmation and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, 4 (3), 343–346.

Krauss, R. M., & Weinheimer, S. (1967). Effects of referent similarity and communication mode on verbal encoding. *Journal of Verbal Learning and Verbal Behavior*, 6, 359-363.

Lazar J, Feng J, Hochheiser H (2010) *Research Methods in Human-Computer Interaction*. Chichester, UK: John Wiley and Sons.

Leudar, I., Thomas, P. and Johnston, M. (1992) 'Self repair in dialogues of schizophrenics: effects of hallucinations and negative symptoms', *Brain and Language*, 43: 487-511.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 14-104.

Levelt, W. J. M. (1993). *Lexical access in speech production*. Cambridge: Blackwell.

Levinson, S. (1983). *Pragmatics*. Cambridge University Press.

Louwerse, M.M., Jeuniaux, P., Hoque, M.E., Wu, J., Lewis, G. (2007). Multimodal communication in computer-mediated map task scenarios. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Mann, W. C. (2002). Dialogue macrogame theory. In *Proceedings of the Third SIGdial workshop, 2002*.

Markus, M. L. (1994). Electronic mail as a medium of managerial choice. *Organization Science*, 4, 5, 502-527.

McCabe, R., Leudar, I. and Healey, P. G. T. (2005). What do you think I think? Theory of Mind and Schizophrenia. In *Proceedings of COGSCI 2005, XXVII Annual Conference of the Cognitive Science Society, July 21-23 2005, Stresa, Italy*.

McGrath, J. E. (1984). *Groups: interaction and performance*. Englewood Cliffs, NJ: Prentice Hall.

McGrath, J. E. (1990). Time matters in groups. In J. Galegher, R. E. Kraut and C. Egido (Eds.), *Intellectual teamwork: Social and technological foundations of cooperative work*, pp. 23-61. Hillsdale, NJ: Erlbaum.

McGrath, J. E. (1991). Time, Interaction and Performance (TIP): A theory of groups. *Small Group Research*, 22, 2, 147-174.

McGrath, J. E. and Hollingshead, A. B. (1993). Putting the 'group' back in group support systems: some theoretical issues about dynamic processes in groups with technological enhancements. In L. M. Jessup and J. S. Valacich (Eds.), *Group Support Systems: New Perspectives*, pp. 78-96. NY: Macmillan.

Mills, G. J. (2007) *Semantic co-ordination in dialogue: the role of direct interaction*. Unpublished PhD Thesis, University of London.

Mills, G. J. (2011) The emergence of procedural conventions in dialogue In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. Boston USA.

Monk, A. (2009). *Common ground in electronically mediated communication*. Morgan & Claypool.

Monk, A. and Gale, C. (2002). A look is worth a thousand words: Full gaze awareness in video-mediated conversation. *Discourse Processes*, 33(3), 257-278.

Morgan, J. L. (1973), Sentence fragments and the notion 'sentence', in 'Issues in Linguistics: Essays in honour of Henry and Rene Kahane', UIP, Urbana.

Morley, I. E. and Stephenson, G. M. (1969). Interpersonal and interparty exchange: A laboratory simulation of an industrial negotiation at the plant level. *British Journal of Psychology*, 60, 543-545.

Nakatani, C., & Hirschberg, J. (1994). A corpus-based study of repair cues in spontaneous speech. *Journal of the Acoustical Society of America*, 95, 1603-1616.

Newlands, A., Anderson, A. H., Mullin, J., & Fleming, A.-M. (1997). Processes of

Collaboration and Communication in Desktop Videoconferencing : Do They Differ From Face-to-Face Interactions ? *Design*.

Newlands, A., Anderson, A. H., Mullin, J., Fleming, A-M. (2000). Processes of Collaboration and Communication in Desktop Videoconferencing: Do They Differ From Face-to-Face Interactions? GOTALOG 2000. *Gothenburg Papers in Computational Linguistics 00-5*.

Newlands, A, Anderson,A. H., Mullin, J. (2003). Adapting communicative strategies to computer-mediated communication: an analysis of task performance and dialogue structure. *Applied Cognitive Psychology, 17, 3, 325–348*.

Nicholson, H. B. M. (2007). *Disfluency in dialogue: Attention, Structure and Function*. Unpublished PhD thesis, University of Edinburgh.

O'Conaill, B. and Whittaker, S. (1997). Characterizing, predicting, and measuring video-mediated communication: a conversational approach. In K. Finn, A. Sellen and S. Wilbur (Eds.), *Video Mediated Communication*, pp. 107-131. Mahwah, NJ: Lawrence Erlbaum.

O'Conaill, B., Whittaker, S. and Wilbur, S. (1993). Conversations over video conferences: An evaluation of the spoken aspects of video-mediated communication. *Human-Computer Interaction, 8, 389-428*.

Orne, M.T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17, 776-783*.

Oviatt, S. and Cohen, P. (1991). Discourse structure and performance efficiency in interactive and non-interactive spoken modalities. *Computer Speech and Language, 5, 297-326*.

Postmes, T., Spears, R., & Lea, M. (2000). The emergence and development of group norms in computer-mediated communication. *Human Communication Research, 26*, 341-371.

Premack, D., & Woodruff, G. (1978). Does the Chimpanzee Have a Theory of Mind. *Behavioral and Brain Sciences, 1(4)*, 515-526.

Purver, M. (2004). *The Theory and Use of Clarification Requests in Dialogue*. Unpublished PhD Thesis, University of London, 2004.

Ross, J. R. (1969). Guess who? In Robert Binnick, Alice Davison, Georgia Green, and Jerry Morgan (eds.), *Papers from the 5th regional meeting of the Chicago Linguistic Society*, 252-286. Chicago Linguistic Society: Chicago, Ill.

Rutter, D. R and Stephenson, G. M. (1975). The role of visual communication in synchronizing communication. *European Journal of Social Psychology, 7*, 29-37.

Sacks, H., Schegloff, E. and Jefferson, G. (1974). A simplest systematics for the organization of turn taking in dialogue. *Language, 50*, 325-345.

Sansone, C., Sachau, D. A., Weir, C. (1989). Effects of instruction on intrinsic interest: The importance of context. *Journal of Personality and Social Psychology, Vol 57(5)*, Nov 1989, 819-829.

Schegloff, E. (1980). What type of Interaction is it to be? *The 18th Annual Meeting of the Association for Computational Linguistics and Parasession on Topics in Interactive Discourse: Proceedings of the Conference*, pp. 81-82.

Schegloff, E. (1982). Discourse as an interactional achievement: some uses of "uh huh" and other things that come between sentences. In D. Tannen (Ed.) *Analyzing Discourse: Text and Talk*. Georgetown University Roundtable on Languages and Linguistics 1981, pp. 71-93. Washington, DC: Georgetown University Press.

Schegloff, E. (1992). Repair after the next turn: the last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology*, 97(5), 1295-1345.

Schegloff, E. (1993). Reflections on quantification in the study of conversation. *Research on Language and Social Interaction*, 26, 99-128.

Schegloff, E. (1997). Third Turn Repair. In G. R. Guy, C. Feagin, D. Schiffrin and J. Baugh (eds.), *Towards a Social Science of Language: Papers in honor of William Labov. Volume 2: Social Interaction and Discourse Structures* (Amsterdam: John Benjamins, 1997), 31-40.

Schegloff, E. (2000). When 'others' initiate repair. *Applied Linguistics*, 21(2), 205-243.

Schegloff, E. (2010). Some other 'Uhm's. *Discourse Processes*, 47:130-174.

Schegloff, E., Jefferson, G. and Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53, 361-382.

Schlangen, D. (2004). *A Coherence-Based Approach to the Interpretation of Non-Sentential Utterances in Dialogue*. Unpublished PhD thesis, University of Edinburgh, Scotland.

Schlangen, D. & Lascarides, A. (2003), The interpretation of non-sentential utterances in dialogue. In A. Rudnicky (ed), *'Proceedings of the 4th SIGdial workshop on Discourse and Dialogue'*, Sapporo, Japan.

Scrivener, SAR., Urquijo, SP. & Palmen, HK. (1996) The use of breakdown analysis in synchronous CSCW system design. In P. Thomas (Ed) *CSCW Requirements and Evaluation*, London, Springer, pp157-172.

Sellen, A. (1992). Speech patterns in video-mediated communication. In *Proceedings of CHI'92 Human Factors in Computing Systems*, 49-59. New York: ACM Press.

Sellen, A. (1995). Remote conversations: the effects of mediating talk with technology. *Human Computer Interaction*, 10, 401-441.

Shadbolt, R.N. (1984) *Constituting Reference in Natural Language Dialogue: The problem of referential opacity*. Unpublished PhD thesis, University of Edinburgh.

Shannon, C. E. and Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.

Shintel, H. and Keysar, B. (2009), Less Is More: A Minimalist Account of Joint Action in Communication. *Topics in Cognitive Science*, 1: 260–273.

Short, J., Williams, E. and Christie, B. (1976). *The social psychology of telecommunications*. New York: Wiley.

Shriberg, E. E. (1994). *Preliminaries to a theory of speech disfluencies*. PhD. Thesis. University of California: Berkeley.

Sidnell, J. (2010). *Conversation Analysis: An Introduction*, London: Wiley-Blackwell.

Skantze, G and Hjalmarsson, A. (2010). Towards incremental speech generation in dialogue systems. In *Proceedings of the SIGDIAL 2010 Conference*, pages 1–8, Tokyo, Japan, September 2010. Association for Computational Linguistics.

Straus, S. G. and McGrath, J. E. (1994). Does the medium matter? The interaction of task type and technology on group performance and member reactions. *Journal of*

Applied Psychology, 79, 1, 87-97.

Themistocleous, M., McCabe, R., Rees, N., Hassan, I., Healey, P.G.T., & Priebe, S. (2009). Establishing mutual understanding in interaction: an analysis of conversational repair in psychiatric consultations. *Communication and Medicine* vol. 6, (2) 165–176.

Thompson, B. H. (1980). Linguistic analysis of natural language communication with computers. In *Proceedings of the Eighth International Conference on Computational Linguistics*, p190–201, Tokyo.

Ulichny, P. (1997). The mis-management of misunderstandings in cross-cultural communication. *International Journal of Pragmatics*, 27, 233-246.

Urquijo, S. P., Scrivener, S. A., and Palmén, H. K. 1993. The use of breakdown analysis in synchronous CSCW system design. In *Proceedings of the Third Conference on European Conference on Computer-Supported Cooperative Work* (Milan, Italy, September 13 - 17, 1993). G. de Michelis, C. Simone, and K. Schmidt, Eds. ECSCW. Kluwer Academic Publishers, Norwell, MA, 281-293.

Volden, J. (2004) ‘Conversational repair in speakers with autism spectrum disorder’, *International Journal of Language and Communication Disorders*, 39: 171-89.

Walker, M., and Whittaker, S. (1990). Mixed Initiative in Dialogue: An investigation into discourse segmentation. In *Proceedings 28th Annual Meeting of the Association for Computational Linguistics*.

Whittaker, S. (1995). Rethinking video as a technology for interpersonal communications: theory and design implications. *International Journal of Man-Machine Studies*, 42, pp 501-529.

Whittaker, S. (2002). Theories and Methods in Mediated Communication. In Graesser, A. (Ed.) *The Handbook of Discourse Processes*. MIT Press.

Whittaker, S., Geelhoed, E. and Robinson, E. (1993). Shared Workspaces: how do they work and when are they useful? *International Journal of Man Machine Studies*, 39, 813-842.

Whittaker, S., Jones, Q. and Terveen, L. (2002). Managing long term conversations: conversation and contact management. In *35th Annual Hawaii International Conference on System Sciences*, 427-437.

Whittaker, S. and O'Conaill, B. (1993). Evaluating videoconferencing. In *Companion Proceedings of CHI'93 Human Factors in Computing Systems*. New York: ACM Press.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Cognition*, 31, 183-194.

Williams, S. H. (1996). "Anaphoric reference and ellipsis resolution in a telephone-based spoken language system for accessing email". In *Proceedings of DAARC'96*.

Yoo, Y. and Alavi, M. (2001). Media and Group Cohesion: relative influences on social presence, task participation, and group consensus. *MIS Quarterly*, 3, 25, 371-390.

Appendix A: Glossary of dependent measures terms and abbreviations

Shared Referent Type	Abbreviation	Description
Endophor	END	A pronoun with an identifiable antecedent
Cataphor	CAT	A pronoun used before the referent is specified
Exaphor	EX	A pronoun without an antecedent referring to something in the immediate environment
Vague Anaphor	VA	A pronoun used without an antecedent that does not refer to something in the immediate environment
Polar Answer	POL	Yes or no answers
Acknowledgement	ACK	A signal of understanding
Prompted NSU Answer	PA	A non-sentential answer that was requested
Unprompted NSU Answer	UNPA	A non-sentential answer that was not solicited
Sluice	SLU	A non-sentential question containing a wh- word
Clarification Ellipsis	CE	A sentence fragment that contains a repeat of a previous contribution due to trouble in understanding
Check	CHK	A query if an other participant has understood a contribution
NSU Query	NSUQ	A sentence fragment asking for more information

Shared Referent Type	Abbreviation	Description
Rejection	REJ	A sentence fragment that disputes a previous contribution
Modifier	MOD	A sentence fragment that describes or qualifies a previous contribution
Continuation	CONT	A sentence fragment that continues directly from a previous contribution
Sentential Ellipsis	SENT	A complete sentence that could be extended by including a previously used phrase

Repair or repair initiation type	Abbreviation	Description
Position one, self-initiated Articulation repair	P1A	An immediate self-edit of a contribution that does not change the meaning e.g. repetition.
Position one, self-initiated Formulation repair	P1F	An immediate self-edit that changes the meaning e.g. word replacement.
Position one, self-initiated Transition Space repair	P1TS	A self-edit or amendment to a contribution that takes place after a possible speaker change.
Position two, other – initiation of repair	P2NTRI	A signal that there is trouble in understanding an other's contribution.
Position two, other - repair	P2OR	A proposed revision or correction of an other's contribution.
Position three, other – initiated self - repair	P3OISR	A self-revision or amendment to a speaker's earlier contribution following an other-initiation.
Position three, self – initiated self - repair	P3SISR	A self-revision or amendment to a speaker's earlier contribution without an other-initiation.

Appendix B: Presumed shared context exemplars from the HCRC map task

In order to illustrate the types of phenomena that have been captured by the coding protocol, these exemplars are presented from randomly selected transcripts. The dialogue transcript code identifying which dialogue the given exemplar came from is given for each exemplar.

B1 Anaphora

G: no I don't have **that** (endophor) (q2nc4)

F: **that's** weird. (vague anaphor) (q3nc5)

G: and have you got a parked van marked **there**, ... beneath the camera shop?
(cataphor) (q4ec4)

G: okay you don't have a forge **there** (exaphor) (q1nc6)

These phenomena appear to be easy to identify from the list of pronouns given at the beginning of the protocol. In certain usages a word such as 'there' is not used in an anaphoric manner; e.g. 'There is no God.'

B2 Answers

G: how far are you underneath it?

F: **about half an inch** (prompted nsu answer) (q2nc4)

G: are you ready?

F: **uh-huh**. (polar answer) (q5ec7)

- F: okay which side will I pass it on?
 G: **on the left.** (prompted nsu answer) (q3nc5)
- G: and about three inches from the side
 F: **right okay.** (acknowledge) (q2ec5)

The map task is negotiated through the exchange of information; in order to check that the task is being completed a series of questions and answers is necessary. Non-sentential answers are often used as a shorter way of providing information on grounding; elliptical usage is highly appropriate here as the context is clear from the question. Unprompted NSU answers are less common in the map task due to the task's nature; they were used more often by the giver as they answered a follower's question, typically with a polar answer, and added an extra piece of information.

B3 Questions

- G: right just move straight down from there then... past the haystack ... to eh...
 F: **past the blacksmith?** (nsu question)(q3ec8)
- G: **okay?** (check) (q5ec7)
- G: now you turn to your right... which will take you towards... the outside of the page.
 F: **what the start side of the page?** (sluice) (q5ec7)
- G: mmhmm, but turn left as if you're walking down
 F: **what?** (Sluice) (q5nc1)
- G: okay about halfway between... the diamond mine... ... and the edge of the page... imagine a.. an obstacle, draw a line

F: **halfway between?** (clarification ellipsis) (q2nc4)

Followers asked most questions, as they attempted to clarify their instructions. The exception to this is the check question category, which was generally used by the giver to ensure that the follower was not having problems.

B4 Statements

G: and then turn right... and go along to about the centre of the page... in a straight line.

F: right.

G: ehm ... well **a wee bit past the centre of the page.** (continuation) (q3nc5)

G: oh there's another plane crash down here though

F: **no only one.** (rejection) (q5ec2)

G: and, then you go ahead mm... you should see... an old mill on your left again.

F: **eh no.** (rejection) (q5nc1)

F: right okay

G: okay?

F: **fine.** (modification) (q2ec5)

G: and that's your end

F: **lovely.** (modification) (q3ec7)

B5 Sentential ellipsis

G: right...right your start point is eh at... the... top left corner right

F: **I know.** (q3ec8)

G: and go along to the finish.

F: **which is where?** (q3nc5)

F: slate mountain?

G: uh-huh

F: **I don't have one.** (q5ec2)

Both the statement and sentential ellipsis categories above were relatively low in frequency; it may not have seemed appropriate or necessary for e.g. a modifier to be used, whereas they are slightly more common in naturally occurring dialogue. Continuations are much more frequent for givers; again, this is due to the need to clarify any misunderstandings (see below). The number of turns that did not use any elliptical phenomena was much higher for givers; this may be due to their need to be precise and specific in their instructions. Once a reference to some named thing has been made, the follower is able to elliptically refer to the giver's instruction.

Appendix C: Repair and repair initiation exemplars from the transcripts

In order to illustrate the dialogue phenomena that have been examined, exemplars taken from the transcripts are presented below. Two major types of position one repair are distinguished by the protocol; articulation and formulation. Articulation repairs demonstrate a problem with creating the contribution; typically repeating or stumbling over what they are trying to say. In contrast, formulation repairs change the meaning of what was about to be said, for example by substituting a word. A third type of less frequent repair is captured by the protocol; transition space repair in which a contribution is potentially completed, but altered before it is responded to. In the map task, these are more common to the instruction giver as they attempt to clarify their instructions to the follower. The following examples illustrate these types of repair:

C(1) Q3EC1

P1 Articulation

Follower: which is due we-- ... **due west?**

C(2) Q1EC6

P1 Formulation

Giver: so you're underneath them ... **between them**

C(3) Q1EC4

P1 Transition Space

Giver: and, start going down southeast **you go past a pine forest on your right**

In position two, there are four types of repair phenomena identified in the Healey et al (2005) repair protocol. Two types were found to be of very low frequency in the corpus and were not used for quantitative analysis in this study. The first of these,

self-initiated other repair occurs when another speaker completes someone else's contribution when a problem is signalled, for example by saying, 'Oh, what's the word..?' The second type that was not included in the statistical analysis was next-turn repair initiation incomplete; when someone signals that they have not heard or understood a contribution but are ignored. The low frequency of this type found in the corpus may reflect the attention that the participants were paying to the task; this can be seen in the significant positive correlation between followers' P2NTRI and givers' P3OISR showing that when a problem was signalled it was dealt with. The remaining two position two repair phenomena, P2NTRI and P2OIOR, were typically used by the follower to signal a problem or clarify an instruction by the giver:

C(4) Q8EC5

P2NTRI Incomplete

Giver: right, then we're going to curve round to where the bottom of the ravine and the vertical from the start meet

Follower: **which direction, curve round.**

Giver: okay?

C(5) Q1EC6

NTRI partial repeat with question word

Giver: past a forge on your right?

Follower: **past a what?**

C(6) Q3NC5

NTRI partial repeat with question word

Follower: so are you going down the way the way I would normally write a "u"?

Giver: **am I what?**

C(7) Q1EC4

NTRI repeat

Follower: dow--, down below that?(P2NTRI)

Giver: up above it. (P3OISR)

Follower: **up above it?** (P2NTRI)

C(8) Q3NC5

NTRI repeat

Giver: so ehm ... I want you to do a "u" round it

Follower: **do a "u" round it?**

C(9) Q1EC6

NTRI question then repeat

Follower: underneath what wheatfields, so i go straight?

Giver: not as far down as that

Follower: **not as far down as that?**

C(10) Q1EC6

NTRI question

Giver: right, that's the end

Follower: **above or below?**

Giver: at the side

C(11) Q1EC8

P2 interpretation/correction

Giver: right to the very end of ... paper

Follower: **the very end of the map?**

C(12) Q3EC1

P2NTRI – P3OISR

Follower: so you want me to go ... **east ... then south?**

Giver: **no, south then east**, we may have a different map

C(13) Q7EC1

P2 repeat then correction

Giver: then you hang a left.

Follower: **left?**

Giver: sorry east ... go east

Follower: **you mean right?**

As seen above, P2NTRI was typically followed by a position three other-initiated repair (P3OISR). The final repair type is position three self-initiated self repair; in this case most commonly when the giver either realised that they had made a mistake and attempt to rephrase or repeat something, or to correct the follower's interpretation of their contribution:

C(14) Q3NC5**P3 (third turn)**

Follower: on the right side of it?

Giver: mm?

Follower: **or the left side of it?** (P3SISR)

C(15) Q1EC8**P3 self-correction**

Giver: right ... now, have you got the hot wells?

Follower: they're over a bit

Giver: **or hot springs?** (P3SISR)

C(16) Q3EC1**P3 rephrase**

Giver: (...) you're paral-- ... parallel with the gallows ... right?

Follower: mmhmm.

Giver: **you're on a level with the gallows?**

Appendix D: Confusion matrices

Confusion Matrix: REPAIR and INITIATIONS (unpublished psychiatric dialogues; two naïve coders)

	P1A	P1F	P1TS	P2NTRI	P2OIOR	P3SISR	P3OISR	none	Total
P1A	283	6	1					15	305
P1F	28	274	6					69	377
P1TS	2	4	4					12	22
P2NTRI				5				2	7
P2OIOR				2	6			7	15
P3SISR						34	1	13	48
P3OISR					1	1	2	1	5
None	13	54	9	4	3	30	8	828	949
Total	326	338	20	11	10	65	11	947	1728
F(e)	57.54	73.74	0.26	0.05	0.09	1.8	0.03	520.08	

Sigma f(e) = 653.59 sigma f(o) = 1436 kappa = 0.73 over 1728 comparisons

Confusion Matrix: SHARED REFERENTS (HCRC map task dialogue Q1EC1; two non-naïve coders)

	END	CAT	EX	VA	POL	PA	UNPA	ACK	SLU	CE	CHK	NSUQ	REJ	MOD	CONT	SENT	NONE	TOTAL
END	119		5	4													3	131
CAT	1	2																3
EX	1		8															9
VA	8	3	8	33													1	53
POL					113			5									2	120
PA						1												1
UNPA							7	1										8
ACK					5			78					2				8	93
SLU									2									2
CE										7								7
CHK											20	1						22
NSUQ							1			1		27			1		2	32
REJ													2				1	3
MOD														1				1
CONT							2					1			13			16
SENT																13	10	23
NONE			4	2		1	2	3	2	1			2		2	3	78	100
TOTAL	129	5	25	39	118	2	12	87	4	9	20	29	6	1	16	16	106	624
F(e)	27.08	.02	.36	3.31	22.69	.003	.15	12.97	.01	.1	.71	1.49	.03	.002	.41	.59	16.99	

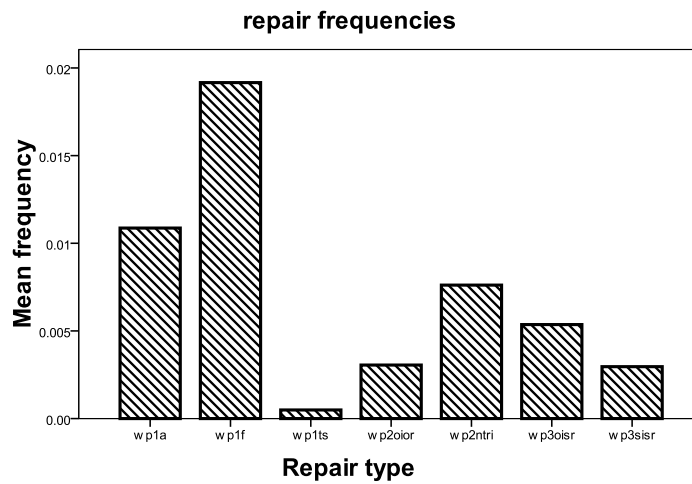
Sigma f(e) = 86.92 sigma f(o) = 524 kappa = 0.81

Confusion Matrix: REPAIR and INITIATIONS (BNC random sample; two non-naïve coders)

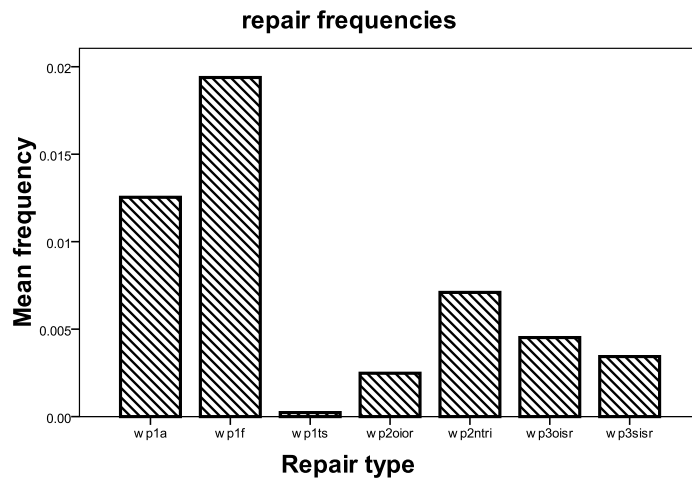
	PIA	PIF	PITS	P2OIOR	P2NTRI	P3OISR	P3SISR	none	total
PIA	7	1							8
PIF		13							13
PITS			3						3
P2OIOR				1					1
P2NTRI					3				3
P3OISR						2			2
P3SISR							5		5
none	1	13	1		2	1		144	162
total	8	27	4	1	5	3	5	144	197
F(e)	0.33	1.78	0.06	0.01	0.08	0.03	0.13	118.42	
Sigma f̂(e) = 120.84 sigma f̂(o) = 178									
Kappa = 0.75									

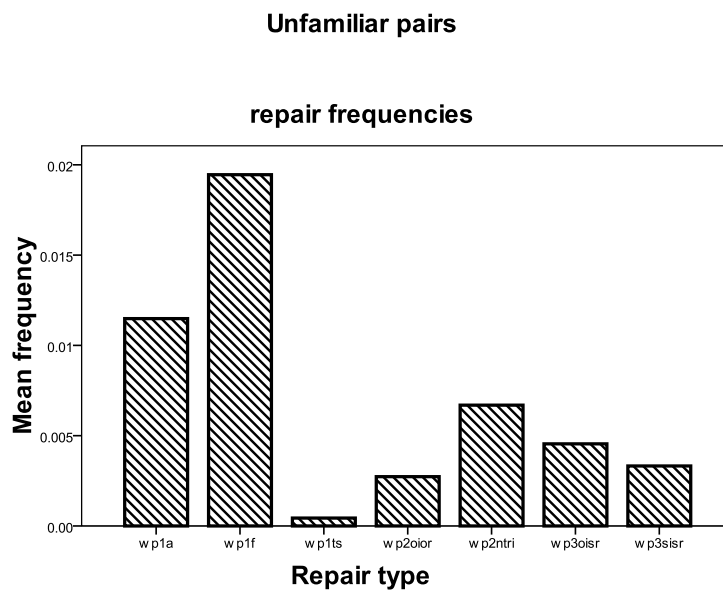
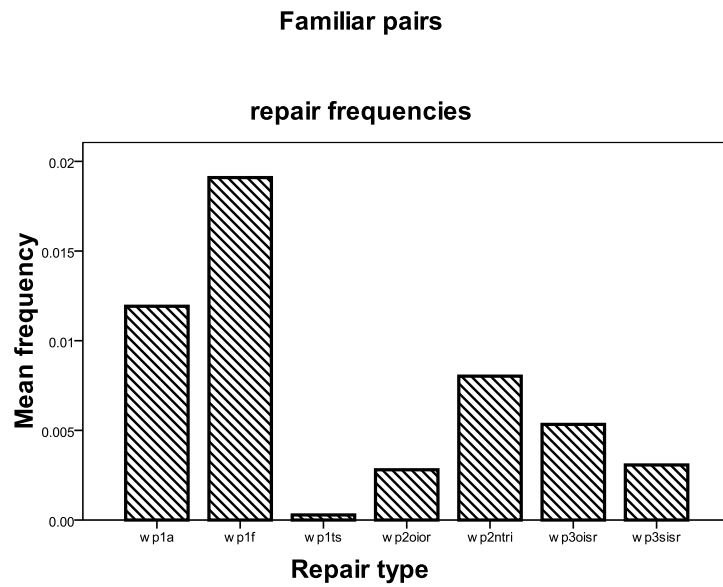
Appendix E: Medium and familiarity repair and shared referents (ellipsis) distributions

Face-to-face

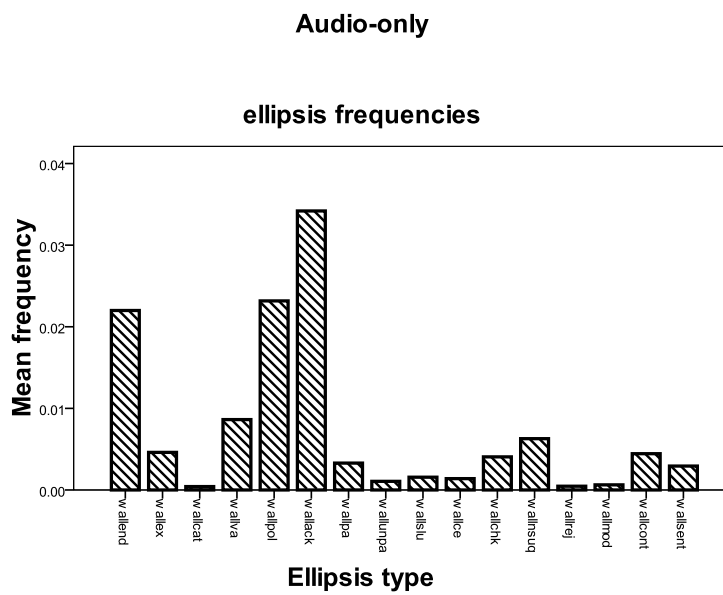
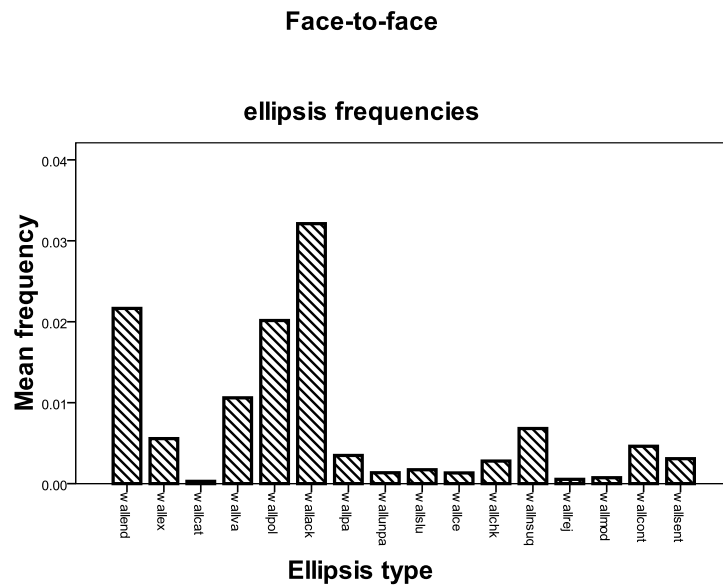


Audio-only





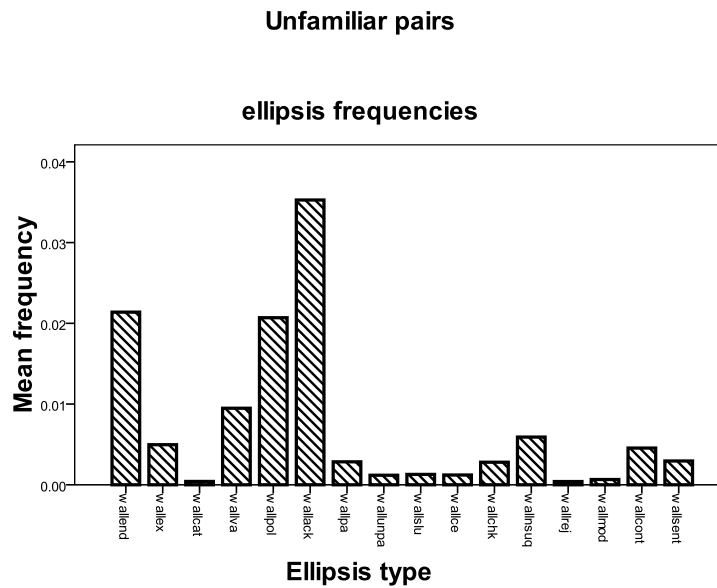
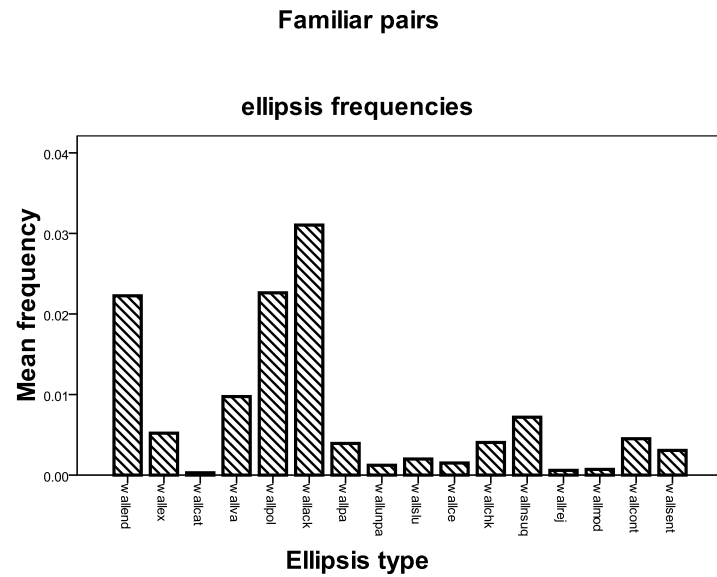
The above figures show the relative distribution of repair phenomena over the factors of medium and familiarity in the HCRC map task corpus. It can be seen that the repair profiles are not varying over the conditions.



The figures here show the relative distributions for ellipsis phenomena in the two medium conditions. The similarities are apparent; medium does not seem to be a factor in this case. This is illustrated by the table below.

ellipsis	Audio		FTF		U	P
	mean	sd	mean	sd		
End	29.95	22.39	23.91	12.83	1787.5	.214
Cat	0.67	1.10	0.38	0.79	1762.5	.096
Ex	6.77	6.18	6.45	6.92	1984	.76
Va	13.11	13.08	12.81	12.06	1945.5	.625
Polar	31.02	21.33	22.64	13.08	1568	.022
Ack	40.61	17.38	33.05	13.72	1503	.009
Pa	3.66	2.55	3.89	3.15	2037	.958
Unpa	1.48	2.17	1.53	1.46	1851	.331
Slu	2.05	2.21	2.03	2.61	1991	.781
Ce	1.80	2.25	1.44	1.71	1947	.618
Chk	5.75	5.90	3.16	4.62	1477.5	.006
Nsuq	8.31	6.02	7.39	4.90	1909.5	.508
Rej	0.72	1.20	0.59	0.75	1999	.793
Mod	0.91	1.99	0.75	1.36	1924.5	.511
Cont	5.31	3.70	4.72	3.47	1851	.346
Sent	3.77	3.13	3.31	3.47	1767.5	.177
none	39.03	21.85	33.11	19.58	1730	.129

Similar results are found for ellipsis over familiarity conditions:



ellipsis	familiar		unfamiliar		U	P
	mean	sd	mean	sd		
End	29.13	16.71	24.73	19.88	1576	.024
Cat	0.48	0.89	0.56	1.04	1986.5	.72
Ex	7.53	7.25	5.69	5.64	1670	.071
Va	14.23	13.15	11.69	11.84	1814.5	.265
Polar	29.95	18.08	23.70	17.75	1527.5	.013
Ack	37.66	17.05	36.00	15.07	1937.5	.598
Pa	4.70	3.32	2.84	1.92	1304	.000
Unpa	1.61	1.55	1.41	2.10	1786	.196
Slu	2.66	2.68	1.42	1.93	1382	.001
Ce	1.88	2.12	1.36	1.85	1756.5	.15
Chk	5.14	5.18	3.77	5.64	1485.5	.007
Nsuq	9.30	5.61	6.41	5.00	1341.5	.001
Rej	0.78	0.88	0.53	1.10	1582.5	.013
Mod	0.77	1.26	0.89	2.06	2017.5	.871
Cont	5.31	3.67	4.72	3.50	1842	.324
Sent	4.02	3.66	3.06	2.84	1754	.157
none	39.58	24.10	32.56	16.51	1722.5	.121

Appendix F: Indices of presumed shared context: between-dialogue and within-role correlations

Variable 1	Variable 2	Spearman Rho	P
Between dialogues (no distinction between task roles)			
Anaphora	NSU Answers	.645	.000
Anaphora	NSU Questions	.662	.000
Anaphora	NSU Statements	.358	.000
Anaphora	Sentential Ellipsis	.239	.007
NSU Answers	NSU Questions	.746	.000
NSU Answers	NSU Statements	.582	.000
NSU Answers	Sentential Ellipsis	.533	.000
NSU Questions	NSU Statements	.441	.000
NSU Questions	Sentential Ellipsis	.374	.000
NSU Statements	Sentential Ellipsis	.270	.002
Within task roles: FOLLOWER			
Anaphora	NSU Answers	.446	.000
Anaphora	NSU Questions	.567	.000
Anaphora	NSU Statements	.421	.000
Anaphora	Sentential Ellipsis	.254	.004
NSU Answers	NSU Questions	.371	.000
NSU Answers	NSU Statements	.322	.000
NSU Answers	Sentential Ellipsis	.205	.020
NSU Questions	NSU Statements	.248	.005
NSU Questions	Sentential Ellipsis	.365	.000
NSU Statements	Sentential Ellipsis	.322	.000
Within task roles: GIVER			
Anaphora	NSU Answers	.512	.000
Anaphora	NSU Questions	.511	.000
Anaphora	NSU Statements	.169	.056
Anaphora	Sentential Ellipsis	.174	.049
NSU Answers	NSU Questions	.490	.000
NSU Answers	NSU Statements	.373	.000
NSU Answers	Sentential Ellipsis	.409	.000
NSU Questions	NSU Statements	.350	.000
NSU Questions	Sentential Ellipsis	.344	.000
NSU Statements	Sentential Ellipsis	.272	.002

Indices of presumed shared context: giver and follower intercorrelations

Variable 1	Variable 2	Spearman Rho	P
Follower Anaphora	Giver Anaphora	.622	.000
Follower Anaphora	Giver NSU Answers	.692	.000
Follower Anaphora	Giver NSU Questions	.538	.000
Follower Anaphora	Giver NSU Statements	.313	.000
Follower Anaphora	Giver Sentential Ellipsis	.212	.016

Variable 1	Variable 2	Spearman Rho	P
Follower NSU Answers	Giver Anaphora	.497	.000
Follower NSU Answers	Giver NSU Answers	.560	.000
Follower NSU Answers	Giver NSU Questions	.689	.000
Follower NSU Answers	Giver NSU Statements	.540	.000
Follower NSU Answers	Giver Sentential Ellipsis	.500	.000

Variable 1	Variable 2	Spearman Rho	P
Follower NSU Questions	Giver Anaphora	.490	.000
Follower NSU Questions	Giver NSU Answers	.733	.000
Follower NSU Questions	Giver NSU Questions	.421	.000
Follower NSU Questions	Giver NSU Statements	.311	.000
Follower NSU Questions	Giver Sentential Ellipsis	.181	.041

Variable 1	Variable 2	Spearman Rho	P
Follower NSU Statements	Giver Anaphora	.340	.000
Follower NSU Statements	Giver NSU Answers	.414	.000
Follower NSU Statements	Giver NSU Questions	.318	.000
Follower NSU Statements	Giver NSU Statements	.214	.015
Follower NSU Statements	Giver Sentential Ellipsis	.162	.068

Variable 1	Variable 2	Spearman Rho	P
Follower Sentential Ellipsis	Giver Anaphora	.169	.057
Follower Sentential Ellipsis	Giver NSU Answers	.415	.000
Follower Sentential Ellipsis	Giver NSU Questions	.261	.003
Follower Sentential Ellipsis	Giver NSU Statements	.052	.559
Follower Sentential Ellipsis	Giver Sentential Ellipsis	.238	.007

Correlations between individual variables (presumed shared context):

Variable 1	Variable 2	rho	p
Deviation from route	F polar answer	-.252	.004
	F NSU Q	-.198	.025
	G and F polar answers	-.235	.008
	G and F NSU Q	-.193	.029
F NSU Q	G polar	.731	.001
G endophor	F NSU Q	.404	.001
G endophor	F CE	.213	.016
G exaphor	F NSU Q	.232	.008
G exaphor	F CE	.275	.002
G VA	F CE	.268	.002
G VA	F sluice	.339	.001
G VA	F NSU Q	.353	.001
F check	G polar	.282	.001
G check	F polar	.663	.001
F check	G rejection	.263	.003
G check	F rejection	.236	.007
F sluice	G polar	.331	.001
F CE	G polar	.280	.001