



Uncommon ground: the distribution of dialogue contexts

Eshghi, Arash

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author

For additional information about this publication click this link.

<https://qmro.qmul.ac.uk/jspui/handle/123456789/551>

Information about this research object was correct at the time of download; we occasionally make corrections to records, please therefore check the published record when citing. For more information contact scholarlycommunications@qmul.ac.uk

Uncommon ground: the distribution of dialogue contexts

Arash Eshghi

Submitted for the degree of Doctor of Philosophy

Queen Mary, University of London

2009

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 conducted by myself unless otherwise indicated.

Arash Eshghi

London, July 6th 2009

Uncommon ground: the distribution of dialogue contexts

Arash Eshghi

Abstract

Context in dialogue is at once regarded as a set of resources enabling successful interpretation and is altered by such interpretations. A key problem for models of dialogue, then, is to specify how the shared context evolves. However, these models have been developed mainly to account for the way context is built up through direct interaction between pairs of participants. In multi-party dialogue, patterns of direct interaction between participants are often more unevenly distributed. This thesis explores the effects of this characteristic on the development of shared contexts.

A corpus analysis of ellipsis shows that side-participants can reach the same level of grounding as speaker and addressee. Such dialogues result in collective contexts that are not reducible to their component dyadic interactions. It is proposed that this is characteristic of dialogues in which a subgroup of the participants are organised into a party, who act as a unified aggregate to carry the conversation forward. Accordingly, the contextual increments arising from a dialogue move from one party member can affect the party as a whole. Grounding, like turn-taking, can therefore operate between parties rather than individuals. An experimental test of this idea is presented which provides evidence for the practical reality of parties.

Two further experiments explore the impact of party membership on the accessibility of context. The results indicate that participants who, for a stretch of talk, fall inside and outside of the interacting parties, effect divergent contextual increments. This is evidence for the emergence of distinct dialogue contexts in the same conversation.

Finally, it is argued that these findings present significant challenges for how formal models of dialogue deal with individual contributions. In particular, they point to the need for such models to index the resulting contextual increments to specific subsets of the participants.

Submitted for the degree of Doctor of Philosophy

Queen Mary, University of London

2009

Contents

1	Introduction	15
1.1	Outline	18
2	Aspects of Dialogue	22
2.1	Introduction: Why dialogue is unique	22
2.1.1	Dialogue as a systematic, collaborative achievement	23
2.1.2	Rules or Norms?	26
2.2	Dialogue as the accumulation of common ground	28
2.3	The Grounding Model: A sketch	30
2.3.1	Taxonomy of Evidence of Understanding	31
2.3.2	Recursion in the grounding cycle	32
2.3.3	Levels of grounding	32
2.3.4	The contribution model is dyadic	33
2.3.5	Textual Vs. Situational Common Ground	33
2.3.6	Conceptual Pacts and Least Collaborative Effort	34
2.3.7	The role of common ground in utterance production and understanding	36
2.4	Sequential coherence in dialogue	38
2.4.1	Adjacency and Conditional Relevance	38
2.4.2	Insertion Sequences	39
2.5	Repair	40
2.6	Coordination between, and Interrelatedness of the behaviour of interactants	42
2.6.1	Synchrony/Alignment in non-verbal behaviour	42
2.6.2	Linguistic Coordination	44
2.7	Computational/formal accounts of dialogue context	47
2.7.1	A very brief sketch of Ginzburg's KoS	48

2.7.2	David Traum: Grounding and Collaborative Conversational Acts	51
3	Scaling up: Multi-party conversation	54
3.1	Introduction	54
3.2	Participation Framework: Critiques of the dyadic model of communication . . .	54
3.2.1	Footing	55
3.2.2	Production Roles: Decomposing the <i>Speaker</i>	56
3.2.3	Reception Roles: Decomposing the Hearer	57
3.2.4	Frames of reference	59
3.2.5	Problems of Multi-party Conversation as a unit	60
3.2.6	Utterance-event	62
3.2.7	The overarching pragmatic problem	63
3.2.8	Issues in assigning participant role given an utterance-event	63
3.2.9	Participation as a demonstrative social role	66
3.2.10	Assignment of roles is not unilateral	66
3.3	Grounding and Participation Framework	67
3.4	Difference in participant role leads to systematically different patterns of com- mon ground	69
3.4.1	Addressees and Overhearers	70
3.4.2	Side participants and Overhearers:	70
3.4.3	Side participants and Direct Addressees: Lack of direct empirical evidence	71
3.4.4	Small group discussions as dyadic dialogue with overhearers?	72
3.4.5	Do small groups attempt to establish common ground among all the members?	73
3.5	Collaborative responsibilities	73
3.6	The pragmatic force of an utterance-event and the problem of multiple interpre- tations	76
3.7	Turn-taking in multi-party dialogues	77
3.7.1	The Simplest Systematics	78
3.7.2	Some Critiques	79
3.7.3	Small group discussions are comprised of a series of pairwise discussions	80
3.7.4	Effect of Group Size on Turn-taking/Grounding	81

3.7.5	Effects of uneven power distribution on turn-taking/grounding	83
3.8	Alignment and priming in multi-party dialogue	84
3.9	Repair in multi-party dialogue	85
3.10	Structuring of multi-party dialogue	87
3.10.1	Exchanges	87
3.10.2	Multiplicity of Conditionally Relevant Contributions	88
3.10.3	Context re-raising	89
3.11	Transformations to Ginzburg's KoS: A short review of the empirical benchmarks	90
4	Collective states of understanding	93
4.1	Introduction	93
4.1.1	Group discussion as dyadic dialogue with overhearers?	94
4.2	Method	95
4.2.1	The AMI Corpus	95
4.2.2	Side Participants to Strips of Dyadic Talk	96
4.3	Analysis of Ellipsis	96
4.4	Quantifying Ellipsis	98
4.4.1	The Coding Protocol	99
4.4.2	Inter-rater Reliability	99
4.5	Ellipsis Taxonomy	100
4.6	Results and Discussion	102
4.6.1	Segments of type LS_{CD} : Assumptions about SPs	102
4.6.2	Segments of type SP_{CD} : Side participant access to 'communal' common ground	104
4.6.3	Segments of type SP_{NC} : Implications for our claims	108
4.6.4	Context locality/saliency	109
4.7	Conclusion	109
5	Parties in Conversation: Grounding by Proxy	111
5.1	Introduction	111
5.1.1	Grounding by proxy	112
5.1.2	When is grounding by proxy expected to hold?	114

5.1.3	Parties in conversation	114
5.1.4	Projection of parties and Corresponding Configurations of the Participa- tion Framework	116
5.1.5	Ramifications of Grounding by Proxy for Participant Role Distinctions .	119
5.2	Method	120
5.2.1	Task and Materials	120
5.2.2	The DiET Experimental Toolkit	121
5.2.3	Experimental Design	122
5.2.4	Participants	124
5.2.5	Procedure	124
5.3	Hypotheses and Coding	124
5.3.1	Initial Predictions	125
5.3.2	Post-hoc Analyses/predictions	125
5.3.3	Coding	126
5.4	Results	127
5.4.1	Local Compensatory Devices: Some Examples	127
5.4.2	Descriptive Statistics	128
5.4.3	Basic Contrasts of Normal and Downgrade Conditions	129
5.4.4	Comparing the Effects of Roles	131
5.5	Discussion	133
5.5.1	Failure to Ground by One Duo Depresses the Level of Grounding by the Duo Team as a Whole	134
5.5.2	Patterns of Direct Interaction	135
5.5.3	The Duo Act and See Themselves as a Party	136
5.5.4	Solo's Compensation: Irreducible Collective States	137
5.6	Conclusion	137
6	Distinguishing Dialogue Contexts	140
6.1	Introduction	140
6.1.1	Ellipsis as window on shared context	141
6.1.2	Accessibility of the Antecedent	143
6.2	Method	144

6.2.1	Task and Materials	144
6.2.2	Experimental Design	146
6.2.3	Participants	146
6.2.4	Procedure	147
6.3	Results	147
6.4	Discussion	149
6.4.1	Strong tendency to interpret a fragment against local context	149
6.4.2	Momentum of dyads	149
6.4.3	Emergence of distinct dialogue contexts	150
6.5	Conclusion	151
7	Collective Vs. Dyadic Contexts	153
7.1	Introduction	153
7.1.1	‘Topic’ as indexed to the participants?	155
7.2	Method	156
7.2.1	Task and Materials	156
7.2.2	Probe Reprise Fragments	158
7.2.3	Experimental Design	158
7.2.4	Participants	159
7.2.5	Procedure	159
7.3	Results	159
7.3.1	Effects of Antecedent and Intervening Context on who responds	160
7.3.2	Effects of Antecedent and Intervening Context on Accessibility of Antecedents	161
7.3.3	Activity Levels of Participants over Collective Vs. Dyadic Items	162
7.3.4	Comparison with the Same-Different Experiment	162
7.3.5	Collective Vs. Dyadic-Different Intervening Contexts	166
7.4	Discussion	167
7.4.1	The directors act as a <i>party</i> on collective items	167
7.4.2	Sensitivity to who said what	170
7.4.3	New Parties Established: Divergence of dialogue contexts	170
7.5	Conclusion	172

8	Concluding Discussions: Ramifications for models of dialogue and future work	174
8.1	Introduction	174
8.2	Empirical Findings	174
8.2.1	Parties in conversation	174
8.2.2	Emergence of distinct Dialogue Contexts	177
8.3	Two systems of Participant Roles: Party Membership vs. Goffman’s Participa- tion Framework	178
8.4	Ginzburg’s KoS	181
8.4.1	Divergence of DGBs	182
8.4.2	Conversational Update Rules could operate on <i>parties</i> in the first instance	183
8.4.3	Pragmatic Pluralism in Dialogue: Relativising update rules to participant role	183
8.5	Contextual updates for Primary vs. Peripheral participants	184
8.5.1	The Cautiously Optimistic Grounding Strategy	184
8.5.2	Primary Co-participants and an essentially defeasible Communal Accep- tance	185
8.5.3	Contextual updates for Peripheral Participants: Systematic Divergence of DGBs	189
8.5.4	Multiple DGBs?	191
8.6	Summary	192
8.7	Open Problems and Future Work	194
8.7.1	Do participants keep track of the distinct dialogue contexts that emerge? .	194
8.7.2	What is the essence of a party and how/why do they emerge?	195
8.7.3	Formal treatment	195
	References	197

List of Figures

4.1	Anaphora Decision Chart	100
-----	-----------------------------------	-----

4.2	Answers Decision Chart	100
4.3	Questions Decision Chart	101
4.4	Statements Decision Chart	101
5.1	Duo 1 Maze	122
5.2	Duo 2 Maze	122
5.3	Solo Maze	122
5.4	Duo 1 Chat Client	123
5.5	Duo 2 Chat Client	123
5.6	Solo Chat Client	123
6.1	Directors' view	144
6.2	Matcher's view	145
6.3	Chat Display	145
6.4	Distribution of responses across conditions	148
6.5	Distribution of the Directors' Interpretation of Probe Reprise Fragments Across Conditions	149
7.1	Director 1's view	157
7.2	Director 2's view	157
7.3	Distribution of All Responses	161
7.4	Interpretations: Collective Vs. 'Dyadic-Different' Intervening Context	167

List of Tables

4.1	Dyadic Segment Type Counts	103
4.2	Ellipses employed by <i>SPs</i> terminating <i>SP_{CD}</i> segments compared to the baseline (BNC)	105
5.1	Descriptive Statistics	129
5.2	Explicit Acknowledgements/Grounding Cues	130

5.3	Frequency of Explicit Address	131
5.4	Explicit Acknowledgements/Grounding Cues	132
5.5	Use of Question Marks	132
5.6	Explicit Addressees By Role	133
6.1	Distribution of the Directors' Interpretation of Probe Reprise Fragments Across Conditions	149
7.1	Distribution of Interpretations by Owners and Non-owners	162
7.2	Distribution of Interpretations when the Owner of Intervening Item is the Same as or Different from the Antecedent Owner	164
7.3	Distribution of Local and Distant interpretations in All the responses received in the 2 experiments	165
7.4	Distribution of Split Utterances Over Collective vs. Dyadic Items	169

List of Dialogue Excerpts

2.1	24
2.2	25
2.3	30
2.4	39
2.5	41
2.6	49
3.1	57
3.2	61
3.3	62
3.4	69
3.5	75
3.6	77
3.7	88

4.1 103
4.2 104
4.3 106
4.4 107
4.5 107
4.6 108
5.1 113
5.2 118
5.3 118
5.4 128
5.5 128
5.6 128
5.7 128
6.1 142
6.2 143
6.3 146

Acknowledgements

Thanking one's supervisor may appear trivial. Well, this one is far from it and I won't be able to do justice to it. Pat Healey has been/is an incredible mentor and friend. Throughout the course of the past four years, he has provided constant help, support and encouragement both on academic and personal grounds. He has had to put up with my more than occasional grumblings about where my thesis was going and whether I was good enough! He has taught me how to think critically. Most importantly he has introduced me to the study of human interaction, which I can say has become a passion; And all of this implicitly in such a way that I always felt as if I was his peer.

I would also like to thank Jonathan Ginzburg who has provided very helpful advice particularly on how to be more concrete about the nature of the predictions that I was trying to make. And of course, my thesis, like many others, draws heavily on his intuitions about context. These have without a doubt acted as a motivating force for some the studies carried out in this thesis.

It is a privilege to have come to know Ruth Kempson, who has on many occasions, including my MPhil viva, provided a substantial set of very constructive criticisms while making certain that I didn't feel bad afterwards! This is the sign of nothing but great care and kindness.

Thanks also go to Graham White, my second supervisor, whose amazingly broad knowledge of Mathematics and Philosophy has always served as a helpful guide.

As for my friends, I would first of all like to thank Gregory Mills for his constant support in the further development of the DiET chat tool, helpful criticism, kindness and most of all for saving me from my occasional delusion that somebody or other had found a definitive answer to the question of meaning! Many thanks go to Matthew Purver for our recent discussions of the intricacies of formal models, and for comments on the last chapter of this thesis. I would also like to express my gratitude to Christine Howes for a meticulous proof reading of the thesis. There were mistakes which I would never have noticed. She has certainly saved me from some embarrassment! Thanks also go to my very dear friend Soroush R. Rad, with whom I have had many heated debates about the Philosophy of Language and Mathematics, some of which have

surely affected the development of the ideas presented here.

Most of all I would like to thank my partner Raha, who has made my endeavours meaningful and my brother Siavash. They have provided the warm and loving environment which I call home.

I'm dedicating this thesis to my parents who have facilitated the little that I have achieved, and without whose unwavering support, past and present, for whatever madness that I've been up to, I would not be where I am.

Chapter 1

Introduction

Studies of the different aspects of human dialogue and interaction have until recently, been mostly limited to dyadic dialogue (interactions between 2 persons)¹. In other words, dyadic dialogue has tended to be treated as the canonical form of interaction/communication. Levinson (1988) argues that the success of the research on dyadic interaction has been at the price of greatly limiting the types of social situations being considered. This is true: taking into account all dialogues transcribed in the British National Corpus (BNC), over 80% involve more than 2 people present according to the BNC transcript headers. This figure is striking and clearly demonstrates the importance of understanding the processes involved in multi-party dialogues.

Indeed, dyadic dialogue is by now a flourishing domain with an ever increasing corpus-based, experimental, theoretical and implementational literature. However, the question of whether multi-party dialogue requires a *unique* treatment remains an open one. To what extent do existing theories of dyadic dialogue (e.g. the psycholinguistic, formal and computational flavours) scale up to multi-party dialogues? Is engaging in a multi-party dialogue a qualitatively different experience? Are the mechanisms posited for dyadic dialogue also at work in multi-party dialogues? If not, which ones are and which ones aren't? Can dyadic dialogue be regarded as a special case of multi-party dialogue? This thesis is a step towards answering some of these general questions, with the specific focus to be narrowed down below.

Many would agree that the primary effect of engaging in dialogue is the accumulation and

¹It may be worth noting that “dialogue” is derived from ‘dia-’ meaning ‘through’ and not ‘di-’ meaning ‘two’ and so covers interactions between multiple participants.

development of shared context or *common ground* (Stalnaker, 1978; Clark & Schaefer, 1989). This is conceived of as a set of resources enabling the successful interpretation of utterances and is altered by such interpretations. It includes but is not exhausted by a dynamic set of assumptions or presuppositions that the interlocutors hold in common at any given time, as part of the background of the conversation. The question of how the shared context evolves on a moment by moment basis through the course of a conversation has received wide attention in psycholinguistic (e.g. (Clark & Schaefer, 1989)), formal (e.g. (Ginzburg, 1996)) and philosophical (e.g. (Lewis, 1979)) circles alike, although as noted, they have tended to concentrate mainly on dyadic dialogue.

It has been argued by psycholinguists that it is only in the aftermath of successful mutual understanding that the participants add the content of what was uttered to the common ground (Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986), and therefore the question of how common ground develops reduces to how people come to establish that something uttered has been understood, on a moment by moment, utterance by utterance basis, i.e. how *grounding* is achieved. Although psycholinguistic theories differ on the mechanisms they posit as responsible for successful communication/mutual understanding, they all emphasise the role of *direct interaction* in this process. In other words, adding to the common ground is not an autonomous mechanism. Participants collaborate directly to achieve grounding, by providing moment by moment feedback, positive or negative, as to whether something said has been understood, and if so, whether they agree with it. With few exceptions, these theories have mainly concentrated on how shared context is built up through direct interaction between *pairs* of participants.

But multi-party dialogue is, unlike dyadic dialogue, conditioned by an overall lack of balance in the participants' levels of direct interaction and involvement with one another. This means that there are differences in the levels of feedback that they provide as to whether and to what extent they are following what is being said and whether they agree with it. It is not only possible but likely that a participant will be left out temporarily or during a whole conversation/topic. Studies have shown that multi-party dialogues are often comprised of a series of pair-wise discussions (e.g. (Parker, 1988)). Moreover, an inspection of the dialogues in the BNC reveals that for each turn, if one looks only at the number of speakers in the preceding 10 turns, about 50% involve only 1 or 2 speakers and 50% more than 2. This is a significant drop from the 80% of all the dialogues, in which there are more than 2 persons present as noted above, suggesting that a great

proportion of naturally occurring multi-party dialogues are comprised of pair-wise interactions.

This immediately raises the question of what, if anything, is the difference between those participants who are in direct interaction for a stretch of talk and those who provide little or no feedback during it. Are the latter at a disadvantage in terms of the level of grounding that they reach of the material they provided little or no feedback on? Studies have shown that they could be (e.g. (Schober & Clark, 1989; Healey & Mills, 2006)). But would these results persist even if this lack of involvement was only ‘temporary’? If not, then what are the differences between a participant who is only momentarily inactive in order to avoid interruptions or overlapping talk, with one who plays a completely passive role? What is the difference between the kinds of common ground increment that they effect in the absence of overt feedback? More broadly, how does shared context evolve and when and for whom is it shared in multi-party dialogue? What are the influential factors in this? Indeed, what should the notion of context look like for multi-party dialogues? These are all open empirical questions which this thesis attempts to address. To this end, this thesis provides a typology of such locally uninvolved third participants, which determines how the various aspects of the shared context of the conversation evolves for them.

First, we show that there are dialogues in which the third silent participant, despite the lack of overt feedback, does reach the same level of grounding as the dyad engaged in direct interaction. These dialogues lead to collective states of understanding or contexts that are not reducible to the component dyadic interactions that gave rise to them (Eshghi & Healey, 2007).

It is then proposed that such collective contexts are the effect, at least, of those dialogues in which subgroups of the participants are organised into *parties* (Schegloff, 1995; Lerner, 1993), such that there are fewer parties than there are individuals. The claim is developed that grounding operates in such dialogues, *in the first instance*, between parties rather than individuals, such that the party acts as a unified aggregate. That is, the party is responsible as a whole, for grounding and carrying the conversation forward and thus satisfying the constraints imposed by the current context of the conversation. In this manner, one member can stand proxy for others in doing this such that it doesn’t matter which member is doing the talking as the rest effect the same contextual increments as that individual. This seems to provide a pragmatic parallel to Schegloff’s (1995) proposal that turn-taking operates in the first instance between parties. An experimental test of this idea is presented which provides causal evidence for the practical reality of parties, in the form of ‘extra’ *intra-party* interactions (what Lerner (1993) has called ‘conferring’), which, all

things being equal, are caused by the failure of one of the party members to ground.

We then develop and test experimentally the claim that in contrast to the dialogues leading to the evenly shared irreducible collective contexts just mentioned, multi-party dialogues can also under some circumstances, give rise to multiple shared contexts that are *unevenly* distributed across the (ratified) participants (Eshghi & Healey, 2009). This is manifested in systematically divergent interpretations of one and the same utterance by different participants owing to the distinct contexts against which the utterance is assessed, a phenomenon we dub ‘Pragmatic Pluralism’ in dialogue. This occurs we argue, when a participant is publicly assumed not to have any concern/responsibility with respect to the current topic of the conversation: she falls outside the interacting parties for a stretch of talk, and as a result is more weakly attached to the local context than the members of the interacting parties.

Finally, it is argued that these results present significant challenges for how formal models of dialogue context deal with individual contributions. In particular, they point to the need for such models to augment the resulting contextual increments with information about which individual/party has processed the contribution and to what degree. Some tentative ideas are put forth at the end of this thesis, as to how this might be achieved in Ginzburg’s (1996, 2009) question based approach to modelling contextual change.

1.1 Outline

Chapters 2 and 3 deal with some of the existing literature on dialogue and interaction.

Chapters 2: Reviews some of the existing literature on the different aspects of dialogue. The aim in this chapter is both to cover the relevant theoretical background to this thesis, and to establish that studies of the different aspects of dialogue have tended to concentrate mainly on dyadic dialogue. These include grounding, conversational structure and adjacency, repair, conceptual pacts, audience design debates, communication accommodation, synchrony and more. At various stages it is pointed out whether/how the notion being introduced, may present further puzzles when it is considered with respect to multi-party dialogue. The discussions in this regard are mostly deferred until the next chapter.

Chapter 3: Aims at a critique of the dyadic speaker/hearer model of communication. It deals with some of the existing literature on multi-party dialogues, with the objective, partly, of bringing to the fore the empirical differences between dyadic and multi-party dialogues and the

open questions and problems that we encounter in scaling up theories of dyadic dialogue. The central notion that we will be concerned with in this chapter is that of *Participation Framework*, constituted by roles other than Speaker and Hearer, an issue which does not really arise in studies of dyadic dialogue. In particular, we will be concerned with the grounds on which such roles are distinguished. Everything else will count towards the characterisation of these roles and the difficulties therein, in terms of patterns of turn-taking, grounding and contextual updates, conversational responsibilities, sequential coherence and coordination mechanisms in general. The particular questions that this thesis will attempt to address are highlighted.

Chapter 4: is a corpus study which uses an analysis of the different surface forms of context-dependency as an index of mutual understanding, showing that in certain multi-party dialogue settings (characterised in the next chapter), active participants of the moment (Speaker and Addressee) assume that a third silent party (a side-participant) has understood/grounded what they're saying to each other, despite the third participant's lack of overt feedback. We also show that the side-participant can utilise, via the use of elliptical expressions, elements introduced into the shared context/common ground that the primary participants are collaboratively building up, in the same way that they themselves do. Withholding of negative feedback by the side participants, then, is taken as positive evidence of understanding. We conclude that these dialogues lead to *collective states of understanding* or collective contexts that, unlike dyadic dialogues, are not reducible to their component dyadic interactions. This leads to the idea, explored experimentally in the next chapter, that a ratified participant can stand proxy for another in grounding/providing evidence of understanding or more broadly in carrying the conversation forward.

Chapter 5: We argue that the emergence of the irreducible collective contexts in the dialogues studied in the previous chapter, is characteristic, at least, of those dialogues in which a subgroup of the ratified participants are organised into a *party* (Schegloff, 1995), such that the party *as a whole* is responsible for carrying the conversation forward. In this manner, a dialogue move from one party member is not associated with that particular individual alone, but the party as a whole, via a process that we dub 'grounding by proxy'. An individual member can thus stand proxy for others in satisfying the constraints imposed by the current context of the conversation. Accordingly, the contextual increments or commitments arising from a dialogue move can, in the absence of overt dissent from that move by other members, affect the whole party, that is, all its

individual members. The joint responsibility of the party members to ground implies that a party member who cedes the floor, bears the responsibility to butt in ‘to compensate’ if anything goes astray; that is, in case of hedgings, pauses, signs of non-understanding and falsities in what is being said by the member who won the floor (the proxy). This is the situation that we try to create experimentally using an online 3-person variant of Garrod and Anderson’s (1987) Maze Game and provide evidence that such mutual responsibilities between party members are operational. We show that this ‘compensation’ often takes the form of ‘extra’ intra-party interactions or what Lerner (1993) has called ‘conferring’ among the party members, which all things being equal, is caused by the apparent failure of one party member to ground. We conclude that, modulo opportunities for intra-party conferring, grounding can operate *in the first instance* between parties rather than individuals.

Chapters 6 and 7: These chapters will be summarised here together as the two experiments carried out are closely intertwined. Given the results from the previous chapter and additional corpus evidence relating to long-distance resolution possibilities for ellipses, we hypothesise that participants who fall inside and outside the locally interacting parties, dubbed as primary and peripheral participants respectively, need to be distinguished. The locally involved/primary participants should be strongly attached to the local conversational context (expressed in terms of the higher saliency of potential antecedents in this context and a strong tendency to interpret distant fragments incorrectly against them), and their level of access to a distant context/antecedent should thus be impaired. In contrast, the locally peripheral participants should have no such attachment to the local context and should therefore find that the distant context, in which they were a primary participant, is more salient/accessible to them. We show experimentally that this is indeed the case: the utterances in the local conversational context (an overarching topic and a set of sub-topics during which the current parties remain unchanged) were available and had priority for the primary participants as antecedents of the (elliptical) utterances they received. By contrast, for a peripheral participant, this context was a ‘distant’ set of utterances, during which they were last (at some earlier stage of the dialogue) a primary participant. In other words, there was a systematic disparity between primary and peripheral participants’ interpretations of ‘*equally elliptical*’ Reprise Fragments, *received from the same individual*, all of which were separated from their antecedents, by the same number of turns/topics on average. We argued that this is evidence for the emergence of distinct dialogue contexts within one and the same conversation, as a result

of one or more previously primary participants falling outside the active/interacting parties for a local stretch of talk/topic. The context that they shared as primary participants with other primary participants, prior to when the new parties were established, is suspended. All the while, the shared context between those who ‘carried on’ as primary participants ‘moves forward’. A single utterance, therefore, can result in divergent contextual updates, applied to primary and peripheral participants’ information states/contexts, resulting in what is dubbed ‘Pragmatic Pluralism’ in multi-party dialogues.

Chapter 8: This chapter rehearses the overall argument of this thesis. It is argued afterwards, that the experimental results obtained in this thesis present significant challenges for formal models of dialogue context. In particular, they indicate that models of dialogue should index the contextual increments arising from an utterance to specific subsets of the participants as these are not guaranteed to be the same for all. Some tentative ideas are presented as to how this might be achieved in Ginzburg’s question based approach to tracking contextual change.

Chapter 2

Aspects of Dialogue

2.1 Introduction: Why dialogue is unique

The study of behaviour while speaking and the study of the behaviour of those who are present to each other but not engaged in talk cannot be analytically separated. (Goffman, 1981)

Most traditional linguistic theories (descendants of Chomskyan Generative Grammars) derive from judgements of acceptability/grammaticality of decontextualised sentences in isolation (Pickering & Garrod, 2004), where the most contextualisation provided is that of a preceding sentence in the form of a conjunction. These are theories of production and comprehension of isolated words and sentences in situations where no interaction is possible:

... the interaction of the speaker and hearer in producing the turn at talk ... has been almost totally ignored in traditional linguistics. (Goodwin, 1981)

Dialogue on the other hand is, *at first glance*, untidy and highly contextualised. According to the tradition, dialogue or natural speech - the most pervasive form of language use and the primary site of language acquisition - is of marginal grammaticality making it an uninteresting and defective data source for the investigation of linguistic structure. Chomsky (1968) suggests that spoken language is:

a highly degenerate sample, in the sense that much of it must be excluded as irrelevant and incorrect - thus the child learns rules of grammar that identify much of

what he has heard as ill-formed, inaccurate, and inappropriate.

The argument, very briefly, is that what people produce in natural conversation often includes fragments, phrasal breaks (syntactically and semantically incomplete), repairs and disfluencies which are taken to be instances of defective *performance*. This makes it extremely difficult to use dialogue as a data source for determining when a sentence is well-formed.

There is, however, a great deal of evidence that dialogue is actually a lot more orderly and systematic than traditional linguistics assumes, hence yielding grammaticality data for dialogue in its own right.

Importantly as we shall also see, dialogue is essentially a shared, collective process, where the talk emerges as a product of the interaction between speakers and hearers. The listeners in dialogue are not passive recipients, rather they are co-creators of the utterances, on a moment by moment basis (Goodwin, 1981; Kraus, 1987; Clark & Schaefer, 1992; Clark, 1996). Moreover, some have argued that production and comprehension are not independent, but coupled processes ((Garrod & Anderson, 1987), but see also Dynamic Syntax (Cann, Kempson, & Marten, 2005) for a grammatical framework which reflects directly the dynamics of incremental growth of interpretations in context and in which production and parsing are seen as basically the same process, hence making it particularly suited to modelling the seamless shift, often mid-utterance, between speaking and hearing in dialogue). This is in direct contrast to the Chomskyan tradition, in which language-production is an autonomous mechanism.

2.1.1 Dialogue as a systematic, collaborative achievement

Work both in the area of Conversational Analysis (a descendent of Ethnomethodology, henceforth CA), and psycholinguistics has shown that if dialogue - as an instance of social interaction - is looked at more carefully, regularities start to emerge. These regularities can often be described as structural norms or conventions governing conversational conduct (we will have something to say about the distinction between rules and norms in the next section).

Having a 'coherent' conversation demands a lot more than what is required in producing or parsing an isolated grammatical sentence. In other words in participating in any social encounter including dialogue, there is a different but dependent set of *competencies* that are required. These competencies involve knowledge of a set of tacit shared norms that the members of a community

adhere to when engaged in talk. The aim of CA is to discover and make these explicit as structural rules. But before we move on to discuss and go into any detail as to what some of these are, a word or two about the CA methodology and a criticism often levied at it are in point.

A common criticism of the CA methodology

The CA methodology draws heavily on retrospective analysis of conversational transcripts. This has always been a point of contention as regards the practice of CA. Others have argued that the analyst has no direct access to people's states of mind during a particular phase of a conversation, at the time of analysis, and hence that the transcripts cannot really count as evidence for participants' orientation to any rule arrived at in such manner. In other words, the criticism is that there is a difference between some rule as merely descriptive of a regularity in a data set, and a rule as prescriptive of certain patterns of behaviour.

Indeed, this would be a valid criticism if it were not for CA's reliance on 'negative' instances where these normative rules are deviated from or breached. These cases are often *marked*: they are accompanied by hesitations, hedgings and often involve participants immediately providing justifications, excuses or accountings for their deviant behaviour. It is precisely such cases which, supposedly, reveal both to the other participants and to the analyst that such norms are in fact operative.

Systematicity in conversation: some examples

Indeed, at first glance, dialogue seems untidy. Nevertheless, there is order in the untidiness. The following examples should demonstrate how speakers use these apparent disfluencies systematically for special purposes, and how hearers on the other hand take them to be meaningful displays of different aspects of a speaker's mental state. This will in turn mean that it would be strange if not wrong, for theoretical linguistics to disregard such phenomena as instances of defective performance:

Goodwin (1980) establishes an intricate connection between what he calls a Restart and the Achievement of Mutual Gaze, in face to face dialogue:

- | | |
|-------|--|
| (2.1) | (1) Debbie: Anyway, (0.2) Uh:, (0.2) We went t- I went to bed. |
| | (2) Chuck: X———— |

Here Debbie restarts her sentence “We went”, to produce the fully grammatical “I went to bed”. She does this exactly when Chuck’s gaze reaches her. This is not coincidental as the pattern here can be found in most such restarts as Goodwin shows. In fact a restart almost always has the following format:

$$(2.2) \left| \begin{array}{ll} (1) & \text{A: [Fragment][Complete Sentence]} \\ (2) & \text{B: X—————} \end{array} \right.$$

The above – as an example of the interaction between speaker and hearer along the course of producing a turn at talk – is highly relevant in the area of language production and hence theoretical linguistics (Goodwin, 1981). Goodwin argues that sentences in everyday dialogue emerge as a product of the interactions between the speaker and hearer (Goodwin, 1979). In other words, speaker and hearer collaborate directly in constructing the turn at talk. The hearer’s behaviour, linguistic or otherwise, has a direct impact on what is produced by the speaker.

Other work in the CA literature regarding interactional structures and norms, includes the organisation of turn-taking, repair, adjacency etc, all of which attempt to spell out the shared methods/devices people use, or the norms people adhere to while engaged in talk. All of these devices require an analysis in terms of the interaction between the speaker and hearer(s) and cannot be otherwise accounted for. Some of these will be expanded on in later sections.

In the psycholinguistic tradition, Clark and Fox Tree (2002) investigate the role of the particles ‘uh’ and ‘um’ in speech, concluding that they have a systematic/purposeful role. Speakers use these fragments to announce that they are initiating what they expect to be a delay in speaking and that they plan this immediately before uttering these fragments. According to the paper, such announcements implicate “that they are searching for a word, are deciding what to say next, want to keep the floor, or want to cede the floor”.

Analogously, Fox Tree and Clark (1997) reveal the systematic role of the alternative pronunciation of the definite article - ‘thi’ instead of the standard ‘thuh’. They find that 81 percent of the instances of the former were followed by a suspension of speech, meaning that people involved were signalling problems at various levels of language production such as articulation and word retrieval.

From a hearer’s perspective, Brennan and Schober (2001) show that listeners comprehend interrupted words better when signals such as ‘uh’s and ‘um’s are present than when they are

absent, or even when there was no interruption at all in the utterance.

The crucial point to be taken here is that there is systematicity to how people manage conversation, despite its apparent disorderliness, and that apparent instances of disfluency often serve various systematic, interactional functions. Moreover, as evident in the Goodwin example above among others, dialogue can only be looked at in terms of an interactive, collaborative achievement by all the parties involved in it.

In general, it seems that we cannot but look at dialogue as a shared, orderly and therefore coordinated activity to which two or more persons contribute, while adhering to a certain set of tacit shared norms. Although much research has addressed these (some of which are summarised in later sections), exactly what these are and indeed the dynamics of how they evolve as a product of interaction, remain open to question.

2.1.2 Rules or Norms?

What we say above seems to imply that the task of the dialogue analyst is, in fact, very similar to the task of the syntactician or semanticist. That is, while the syntactician seeks to describe what it is for a sequence of words to be well-formed, or the semanticist seeks to describe how semantic concepts of various types can be combined to yield propositions, the dialogue analyst strives to make explicit a set of shared norms that people adhere to when conversing, thus seeking to establish a notion of dialogue *coherence*. She tries, in other words, to spell out what it is for a sequence of dialogue utterances to be coherent or well-formed. As we will see, a lot of work in the computational literature on dialogue (e.g. Ginzburg (2009) or Traum (1994) both to be sketched in later sections), drawing on the work by Conversational Analysts and Psycholinguists, attempts to do just this. The result is always a protocol for interaction. Given the current ‘dialogue state’, a protocol essentially constrains the set of next actions or dialogue moves that can ensue. Interaction protocols arise as a result of combining a set of finer-grained interactional norms. As we will see shortly, interaction protocols under Ginzburg’s (2009) KoS, arise as a result of the composition of a set of Conversational Rules. We won’t expand on the details here.

Nevertheless, none of these models seem to adequately address the very complex problem of what happens when such norms are deviated from or breached. This is a significant aspect of dialogue and interaction, and seems to be what makes language and meaning dynamic, normative and emergent phenomena. Let’s explore this a little. Note that no claims are being made here. What follows is only intended as a brief discussion.

A person's deviation from an interactional norm often carries 'extra' meaning or significance. Such meanings have been called *implicatures* by the philosopher Grice (1975), who has tried to give a systematic account of just how these non-surface meanings are generated in virtue of the mutual recognition of that norm being shared and in place. People often intentionally - in the sense explicated by Grice (1957), that is, as a non-natural meaning, which is intended and recognised as such by the hearer - exploit this. A simple example is that greetings are usually followed by greetings (as we will discuss in more detail in later sections this is an instance of the more general Conversational Analytic notion of 'conditional relevance'). But we have all experienced the unpleasant moment when someone does not return our Hello, though we're certain that they've heard us. We would in such cases invariably think that something was wrong, perhaps immediately asking for an account if we couldn't readily make a guess at it.

Nevertheless, Grice formulates such norms quite abstractly (e.g. his Maxim of Relevance), indeed leaving plenty of room for variability in context. But it is precisely when such norms are breached that it is conceivable that there will be a displacement, a joint modification - via coordination between speaker and hearer - to the meaning of the norm itself (e.g. what it is to be relevant in that particular context of use), and with the possibility of such displacement becoming permanent and hence inseparable from the norm, so far as the same pair of people are considered.

Semantically, a parallel argument can be made that the meaning of words or concepts themselves are normative and emergent (we will discuss this in more detail in later sections). Upon repeated use within novel contexts, words can get conventionally associated with concepts different from those before, i.e. semantic ontology is essentially dynamic. Strictly speaking, a dictionary has the shelf life of less than a day! This also implies that, so far as word meaning is concerned, no 2 speakers of a language community, if abstracted away from any particular context of use (a dialogue), would speak exactly the same dialect of that language as they have, one way or the other, had to deal with different situations using the same words. But as we will see, there are devices in dialogue, such as repair, for the local resolution of such mismatches, so that this will not and indeed does not lead to complete mayhem and miscommunication.

What is the status of Grammar in this regard? Can instances of 'ungrammaticality' be intended as meaningful and recognised as such by a hearer? It would certainly be more difficult to construct examples of this sort. However, it does seem possible. For example, one might do this for poetic or rhythmic effect. This happens quite often in traditional Persian poetry where there

are very strict rhythmic rules, and as a result the poet violates the Grammar of the language to accommodate the required rhythm. Some of these instances of ‘ungrammaticality’ may well have been routinised, i.e. they would now be used in everyday conversation, hence rendering them ‘grammatical’. But note also that in order to get a meaningful effect such as the absence of the ‘Hello’ above, the interlocutor would have to know that the person breaching the rule is in fact a fluent speaker of that language and that he’s breaching it for an intended effect. We should note that within the Chomskyan tradition, syntactic rules (or at least the principles in Chomsky’s Universal Grammar) are supposed to reflect fixed rules, characteristic of the brain’s language processing module or indeed of ‘the language of thought’ in general. But what we say above implies that even syntax may be normative, and hence emergent from social interaction and actual language use. There have been arguments to this effect: see e.g. Ford, Fox, and Thompson (2002) for arguments that syntax can be emergent from turn taking pressures. There are also arguments in historical linguistics that diachronic syntactic change can be caused by processing and production pressures.

In any case, this is an extremely interesting, fundamental, intricate and perhaps somewhat dangerous debate (it has a large number of important philosophical consequences), to which many theses could be dedicated. We will not go into any further detail in this regard. We will, however, come back briefly to the discussion of the relationship between Conversational Rules and interaction protocols in the context of Ginzburg’s model of dialogue context.

2.2 Dialogue as the accumulation of common ground

The primary effect of engaging in a dialogue is the accumulation of mutual-knowledge or as we shall prefer *common ground*. Stalnaker (1978) defines common ground as ‘the propositions whose truth he [the speaker] takes for granted as part of the background of the conversation’. Upon the production of an utterance by a dialogue participant, the content should somehow be added to the common ground, since a new utterance would inevitably presuppose the information already conveyed by the previous utterances in the dialogue: ‘the essential effect of an assertion is to change the presuppositions of the participants in the conversation by adding the content of what is asserted to what is presupposed.’ (Stalnaker, 1978). This brings us to the question of what it takes to contribute to a dialogue. We need to know the systematics of the accumulation of common ground through the course of a conversation. A theory of dialogue context is needed.

This is essential for any adequate account of how people are able to participate in dialogue:

- **Comprehension:** We need to interpret utterances. But dialogue is demonstrably a heavily context-dependant form of language use. Utterances are often incomplete sentences (fragments) the resolution of which relies heavily on what is assumed to be already part of the participants' common ground.
- **Production:** We need to produce the right utterance at the right time. But the relevance or felicity of an utterance in any conversation is largely dependant on what the speaker takes to be common ground, between himself and his interlocutors.

Prior to the emergence of the grounding model (Clark & Schaefer, 1989), most models of discourse (e.g. (Kamp, 1981; Lewis, 1979; Stalnaker, 1978)) assumed that the accumulation of common ground is a unilateral mechanism - that contributing to the conversation only requires action (uttering the right words at the right time) from the current speaker, whereby the content of what is uttered is added to the common ground between the speaker and all those in the perceptual range of the utterance. Models of communication such as Searle's speech act theory (Searle, 1969), also assume perfect communication where the speaker has complete control over the communicative process and thus the securing of uptake by the interlocutor is not considered in the model:

[It is] not the symbol or word or sentence or even the token, but rather the production [by the speaker] of the token in the performance of the speech act that constitutes the basic unit of linguistic communication

These models then, all assume that dialogue participants add to the common ground autonomously.

In their seminal paper, Clark and Schaefer (1989) criticise this view, proposing that contributing to discourse is a joint collaborative process. Communication, on this view, is a form of joint action, with conversationalists often likened to dancers or musicians, performing together as members of an ensemble. Having a conversation demands that the participants coordinate both on the content of what they say (e.g. sequential coherence) and the process (e.g. turn-taking). The rest of this section will elaborate on the different aspects of the coordination between participants in the orderly accumulation of common ground.

In conversation, the main issue is that of establishing that something uttered has been understood by the interlocutors. According to Clark and Schaefer (1989); Clark and Brennan (1991) common ground cannot be properly updated without a joint process referred to as *grounding*. Grounding can be roughly defined as the process by which participants mutually establish that a certain utterance has been understood and that the conversation can now move on. The utterance is then said to be grounded.

2.3 The Grounding Model: A sketch

What does it take to contribute to a conversation? Consider the example below:

- (2.3) | (1) Alan: Now, -um, Do you and your husband have a j- car?
 (2) Barbara: have a car?
 (3) Alan: Yeah
 (4) Barbara: Yes
 (5) Alan: What is the make of it?
 (6) Barbara: It's a Ford

Speech act theorists, who conceive of communication as an individualistic process/capacity, would argue that what Alan has done here is ask Barbara whether she and her husband have a car and that in this way he has carried the conversation forward. However this is not quite what happens. What has really happened is that Alan has initially failed to get his question across, which is why Barbara is forced to request clarification by uttering “- have a car?”. It is only after Alan responds to the clarification request that Barbara answers the original question. This shows that asking a question involves more than the production of an interrogative. Before any question - or any utterance type for that matter - can be appropriately responded to, the dialogue participants need to establish mutually that the question has been understood; i.e. Barbara should believe that she has understood the utterance and Alan must be able to believe that Barbara believes that she has understood the question. An utterance needs to be grounded before any relevant response can ensue.

According to Clark and Schaefer (1989), an utterance is grounded 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 referred to as *the grounding criterion*. Grounding then, is the process by which conversational participants try to reach this mutual belief.

So, what does it take to contribute to a conversation? Clark and Schaefer (1989) propose that any complete contribution in dialogue is composed of the following two phases:

1. **Presentation phase:** *A* presents utterance *u* for *B* to consider. He does so on the assumption that if *B*, gives evidence *e* or stronger, he can believe that she understands what he means by *u*.
2. **Acceptance phase:** *B* accepts utterance *u* by giving evidence *e* that she believes she understands what *A* means by *u*. She does so, on the assumption that once *A* registers that evidence, he will also believe that she understands.

Grounding is generally the desired end result of the two phases. The acceptance phase is necessary because when *A* finishes an utterance, *B* can be in any one of the following states:

1. *B* didn't notice that *A* uttered any *u*.
2. *B* noticed that *A* uttered some *u* (but wasn't in state 3).
3. *B* correctly heard *u* (but wasn't in state 4).
4. *B* understood what *A* meant by *u*.

Below in section 2.5, we will consider cases of miscommunication, i.e. when *B* is not in state 3, thereby necessitating meta-communicative interaction or repair, before the conversation can move on.

2.3.1 Taxonomy of Evidence of Understanding

As already implied, the exchange of evidence of understanding (in the acceptance phase) is crucial to the interlocutors reaching the mutual belief that something said is understood. Different forms of evidence of understanding have been analysed and classified in Clark and Schaefer (1989); Clark and Brennan (1991). There are two main categories, namely positive and negative evidence. In dialogue excerpt 2.3, “-have a car?” is an instance of negative evidence or a request for repair (more on this below). However, absence of negative evidence is not enough to move the conversation forward. A speaker also expects positive evidence of understanding from the hearer that their utterance has been understood.

Clark and Brennan (1991) provide the following taxonomy of positive evidence, ordered from strongest to weakest:

1. **Demonstration:** B demonstrates all or part of what he has understood A to mean.
2. **Acknowledgement:** B produces a backchannel response (Schegloff, 1982).
3. **Relevant next turn:** B produces a relevant next turn. E.g. if a question is asked, an appropriate answer is a relevant next turn.
4. **Continued attention:** A non-verbal form, which simply points to positive signs of attention like gaze.

2.3.2 Recursion in the grounding cycle

The grounding cycle, presented above, is recursive in the sense that any acceptance is itself an utterance, and hence a presentation per se, which needs to be accepted before it can take effect (grounding of the initial presentation), meaning that we're faced with the problem of infinite regress. Clark and Schaefer (1989) attempt to rectify this problem by positing the *Strength Of Evidence Principle*:

The participants expect that if evidence e_0 is needed for accepting presentation u_0 , and evidence e_1 for accepting evidence e_0 , then e_1 is weaker than e_0 .

So grounding cycles should end in continued attention, at which point the acceptance in question takes effect and the corresponding presentation is grounded.

2.3.3 Levels of grounding

Analogously to the model developed by Allwood (1995), Clark (1996) distinguishes 4 levels of (collaborative) action in dialogue. These correspond to distinct levels at which an utterance or a presentation u may be grounded. They have different terminologies for these levels. Here, we will adopt Clark's terminology:

- Acceptance: whether the hearer has integrated the content of u
- Recognition: whether the hearer has understood what u means
- Identification: whether the hearer has heard/perceived u

- Attention: whether the speaker and hearer have established a channel of communication

Note that each of these levels presupposes the next. The levels are distinct to the extent that the grounding of any utterance can be negotiated on each level. For example, while a hearer may understand fully what *u* means (Recognition), she might not agree to what is proposed in *u* (Acceptance): “A: That’s a nice building. B: No. I think it’s actually quite ugly”. Or if *u* is a question, the hearer might be unwilling to answer the question, although she has fully understood it: “I’d prefer not to answer that right now”.

So after any *u* is presented, the speaker and hearer are faced with questions concerning each of these levels. It follows that for any speaker and hearer, for each presentation, each of these levels of grounding is represented in the common ground.

Corresponding to each of these levels, there are specific kinds of feedback that the hearer can use to signal whether or not a presentation has been grounded at that particular level. For instance, acknowledgements can be used to signal both Recognition and Acceptance, while rejections (e.g. “no”) can be used to signal Recognition while rejecting the content of an utterance (Acceptance).

In the next chapter, where we discuss the differences between dyadic and multi-party dialogues, we will consider the idea that participants in different roles may or may not have access to feedback ‘tool sets’ corresponding to the above levels, and moreover that participant role plays a part in the assumptions made by a speaker about a participant’s grounding of an utterance on each level.

2.3.4 The contribution model is dyadic

We should note that the grounding/contribution model outlined above - taken here as an interaction protocol for establishing that a presentation has been mutually accepted - is dyadic. It assumes that there are only 2 people engaged in the dialogue, despite the fact that the grounding criterion vaguely covers multi-party settings. It is unclear how the model, as specifying a systematic account of grounding in conversation, would scale up to multi-party dialogues. This is the main theme of this thesis and will be taken up in considerable detail in the next chapter, where we consider different forms of participation and their implications for patterns of grounding.

2.3.5 Textual Vs. Situational Common Ground

Once grounding has been achieved and a particular piece of information has been introduced into the common ground, the participants can refer to it freely and assume that the other will be

able to understand these subsequent references. Moreover, these subsequent references are made using definite reference clauses and anaphora which presuppose that the interlocutor knows about the referent. In example 2.3, once Alan's question is grounded and answered, both participants can subsequently refer to Barbara's car as 'it' or 'the car'. In fact, participants have access to two different aspects of the common ground which together constitute the overall *discourse representation*.

Clark and Schaefer (1992) make the distinction between these two components: *situational* and *textual common ground*. The former refers to the semantic common ground. In other words, this is a set of representations in which all references have been resolved, effectively representing a set of beliefs, so far established as common by the interlocutors. Note that in this component information about the linguistic means used to establish these propositions is not represented.

The textual common ground, comprises a linguistic record, which represents the linguistic means (phonology, words and syntax) that have been used in the dialogue to mutually establish the propositions in the situational common ground. Some of this information will not be retained for long. For example people are not good at remembering past utterances verbatim (see e.g. Marslen-Wilson and Tyler (1976)). But other information such as the way in which a person has conceptualised an entity, using a particular definite description or name - a phrase-concept pairing - may be retained throughout the course of the conversation and sometimes beyond, so that subsequent references to the same entity will employ the same phrase. Moreover, as we shall see below, people are sensitive to what word has been used with a particular partner so that the same term might not be used with others to refer to the same entity.

But what empirical evidence is there that people retain in their common ground such textual representations or word-concept pairings? Or indeed that people remember which term they've used with a particular interlocutor? In other words, is the linguistic common ground indexed to specific dialogue partners? This brings us to what Brennan and Clark (1996) have dubbed *conceptual pacts*.

2.3.6 Conceptual Pacts and Least Collaborative Effort

Clark and Wilkes-Gibbs (1986) and Brennan and Clark (1996) carried out a series of experiments upon pairs of subjects, with one subject assigned to the role of *Director* and the other to the role of *Matcher*. In these experiments the Director had to describe a sequence of abstract figures (so-called Chinese Tangrams) to the Matcher, so that she, who had the same figures at her disposal

but shuffled, could arrange them into the sequence described by the Director.

They found, first, that once a nominal, or a definite phrase had been used by the director with a particular partner, the director tended to re-use that phrase again in subsequent interactions with the same partner, even if the current context/figure meant that the expression was not totally appropriate.

Second, they found that once a term had been used to refer to one of the figures with one partner, if the director then went on in a later trial to describe the same item to a different partner she would often NOT use the same term, and that indeed this would require a brand new grounding cycle for the different term used.

These observations therefore, provide evidence for partner-specific choices of referential expressions, so that in the textual common ground the linguistic expressions used to refer to different entities are somehow indexed to specific participants, a property of utterances which fits under the very general rubric of *audience design*, discussed in more detail below.

In section 2.6 below, we will review some of the literature on the phenomenon of alignment; the interrelatedness of the behaviour of people engaged in interaction at a local level, i.e. the way a speaker's behaviour (linguistic or otherwise) directly influences the immediate subsequent behaviour of another participant - that is, NOT just as a result of grounding a common perspective or a conceptual pact.

A third result recurring in all referential tasks of this sort, is that once an expression has been grounded as referring to a specific figure, in subsequent trials with the same partner, when the director encounters the same figure again, the expression used tends to contract, get shorter than the one used previously. For example, "the dancer with his leg up in the air" could become "the dancer" in later trials. Moreover, the grounding cycle itself gets shorter over repeated trials.

To account for this, Clark and Wilkes-Gibbs (1986), in their model of grounding definite references, propose *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.

Apart from the contraction of referring expressions, the principle of least collaborative effort is supposed to account for people's preference for both self-initiation and self-repairs (Schegloff, 1977). As we shall see in section 2.5 below, both of these result in shorter grounding cycles, and hence require less joint effort.

This principle contrasts with individualistic, speaker-centred accounts of meaning and communication, such as the Gricean maxims of Quantity and Manner, or Searle's principle of least effort, which would predict that once an expression has been used successfully by the speaker, she should carry on using the same expression on subsequent occasions. Clark and Wilkes-Gibbs (1986); Clark and Brennan (1991) trace this problem back to the assumption of perfect communication in such models. "They assume flawless presentations and trouble-free acceptances and therefore cannot do justice to what really happens in conversation" (Clark & Brennan, 1991)

2.3.7 The role of common ground in utterance production and understanding

Audience design is a pervasive property of utterances in both conversational and written form (Bell, 1984). It designates how an addressee has an impact on the speakers' utterance formulations or behaviour in general, i.e. how speakers tailor their utterances to their specific interlocutor(s), thus taking the unique common ground that they have with their interlocutors into account. Gergen (1972) notes in this regard that when he writes letters to close friends, depending on his current knowledge of and relationship with that friend, he comes across as a completely different person in each case: "In one, I was morose, pouring out a philosophy of existential sorrow; in another I was a lusty realist; in a third I was a lighthearted jokester".

Successful communication in Clark's grounding model, in line with the Gricean cooperativity principle (Grice, 1975), relies heavily on speakers cooperatively designing their utterances for the benefit of their addressees, through reasoning about the addressee's state of mind (beliefs, intentions etc.), and hence taking the unique common ground that they have with their addressees into account. The Gricean Maxims in the grounding model, are supposedly instantiated by speakers in specific cases, against the common ground between them and the addressees. For example, what counts as "as informative as required" (Grice's Maxim of Quantity) when addressed to A, may not be so when addressed to B, and this is determined by the common ground.

Some of the empirical evidence for audience design relates to the semantic content of the utterances produced. For example, Fussell and Krauss (1992) showed that the referential expressions that speakers used to refer to photographs, reflected their prior beliefs (possibly mistaken) about their addressees. For example if they thought that the addressee was likely to know about the person in the photograph they tended to use more concise descriptions, or often names, whereas otherwise, they produced more elaborate descriptions. There is also evidence that through the course of a conversation, people can dynamically adjust their behaviour, e.g. the

amount of detail they provide about some referent/concept, as the degree to which the addressee knows about the concept is gradually revealed (Isaacs & Clark, 1987).

Other evidence has to do with the form of speaker's utterances, e.g. phonology, speech-rate and syntax; speakers may speak more slowly or use less complex syntactic constructions if they believe that the addressee is not a fluent speaker of the language. We will discuss such 'accommodations' by the speaker, in a little more detail in section 2.6.2.

So far, we've talked about how common ground guides language production. Common ground also has an impact on how others interpret the speaker's utterances. The addressee may take what he knows about the knowledge/situational representations of the speaker into account when interpreting her utterances (see e.g. (Hannah & Tanenhaus, 2004)). Comprehension, then, is also guided in the grounding model by assumptions of cooperativity. For example, suppose a colleague asks me where I live, and I say, in London (he obviously knows this already). This would probably be taken by him as some sort of joke or insult, as he assumes that I am intentionally flouting Grice's Maxim of Quantity. So what the addressee takes the speaker to know (common ground), guides his interpretation of the speaker's words. Another example is that for a reference by a speaker that is ambiguous for the addressee, she might perform disambiguation by inferring that the speaker could not be talking about some referent because he could not possibly know about it.

Although it is uncontroversial that common ground plays a part in both production and comprehension of utterances, the exact nature of the psychological memory processes which enable access to common ground - and hence a characterisation of the situations in which this may or may not occur - are still a great source of contention. For example, Keysar and colleagues (e.g. (Keysar, 2007; Horton & Gerrig, 2005; Horton & Keysar, 1996)) have been arguing, through a series of studies, that both language production/audience design and interpretation are constrained by subconscious memory processes and that therefore the time available for the retrieval of representations has a direct impact upon the degree to which common ground is taken into account during production/comprehension. These studies suggest that people's real-time utterance production/comprehension in dialogue, are at least initially egocentric, depending upon their own private knowledge/memory representations (which are, they argue, more accessible through quicker, less computationally expensive, 'automatic'/post-conscious mechanisms), and that common ground is only consulted when something goes wrong (e.g. negative feedback from

interlocutor). Nevertheless, this contention does not seem to have been resolved as there are conflicting pieces of empirical evidence in both directions (see Schober and Brennan (2003) for a review).

We should note that issues of audience design and partner specificity in language use are of great importance in multi-party dialogues. It is, as yet, unclear to what degree speakers in multi-party dialogues take all participants in the conversation into account when they plan their utterances. This becomes an extremely complex, if not impossible, task when the participants' knowledge about the topic under discussion is not aligned, e.g. when two participants are experts but one is a novice, or when group size increases, forcing speakers in such circumstances to issue more 'generic' utterances.

With regards to comprehension, the impact of common ground on a participant's interpretation of an utterance implies that in multi-party dialogue, discrepancies between what is mutually known to a speaker and each hearer, can lead to differences in interpretation of the same utterance by different hearers, in one and the same conversation. This latter problem will be of great concern in the final chapters of this thesis. We will come back to these issues once we've introduced multi-party dialogues in the next chapter.

2.4 Sequential coherence in dialogue

Within the CA tradition, a lot of attention has been paid to the sequential coherence of an unfolding interaction. In this section we will take a glance at some of their findings in this regard.

2.4.1 Adjacency and Conditional Relevance

As we saw above, any presentation requires some form of acceptance from the interlocutor. It is very often the case that an acceptance is implicit in an appropriate reaction to the presentation. But what counts as appropriate here?

It turns out that for a given type of utterance, there is indeed a very restricted set of appropriate types of utterance that can be produced as a response. For example, a greeting is followed by a greeting, a question by an answer, while an assertion requires an acceptance or rejection. This is the sense in which a turn is sequentially implicative, meaning that a turn projects a response. This leads to units which are larger than the individual utterance; *the exchange* or *the adjacency pair*:

A given sequence will be composed of an utterance that is a first pair part produced by one speaker directly followed by the production by a different speaker, of an utterance which is (a) a second pair part, and (b) is from the first pair type as the first utterance in the sequence is a member of. (Schegloff & Sacks, 1973)

The production of a first pair part, then, makes the second pair part ‘*conditionally relevant*’:

By conditional relevance of one item on another we mean: given the first, the second is expectable; upon its occurrence it can be seen to be a second item to the first; upon its non-occurrence it can be seen to be officially absent - all this provided by the occurrence of the first item. (Schegloff, 1972b)

Importantly for grounding, such projectability, or sequential implicativeness allows participants to monitor stage by stage, turn by turn, the others’ interpretation of their utterances, so that each turn displays the uptake of the previous one. So if B produces T2 in response to A’s turn T1, this reveals B’s interpretation of T1 to A. Below in section 2.5, we will see how this allows for subsequent *third position repairs* by A.

2.4.2 Insertion Sequences

Often, the second pair part of an adjacency pair cannot be found immediately after the first pair part, but separated from it by another embedded adjacency pair, referred to as an insertion sequence (Schegloff, 1972a), as in example 2.4 below:

- | | | |
|-------|--|-------------------------------------|
| (2.4) | | (1) A: Do you want to go for lunch? |
| | | (2) B: Where are you going? |
| | | (3) A: Turkish shop down the road |
| | | (4) B: Yes. I like that place. |

Insertion sequences then, are preliminary from the point of view of the responder, to the response (second pair part) subsequently produced to the initial first-pair part. As we will see shortly, this is the reason why in computational models of dialogue context, the QUD (Questions Under Discussions) element as modelling the highly constrained ‘discursive potential’ in conversation, is a partially ordered set (often a stack).

2.5 Repair

Whatever the case may be as regards the role of common ground in interpretation, non/mis - understandings can arise, among other things, as a result of speakers making erroneous presuppositions about what they share (linguistically or otherwise in their common ground) with their interlocutor, which can in turn arise because speakers are behaving egocentrically. Keysar (2007) conceives of the latter as a systematic source of miscommunication. Considering moreover that dialogue is spontaneous, this is more likely to occur in dialogue than in written language, where a participant has time for revision. A characteristic feature of conversation, then, is the work that people do to resolve such problems.

So what conversational mechanisms are there for rectifying such problems in understanding? Below, we will summarise the extensive analysis within the CA tradition of different types of *repair* phenomena.

In contrast to models of repair which try to characterise specific kinds of trouble in understanding, the reasons behind them and the corresponding techniques that people use to deal with them (see e.g. Clark and Marshal's (1981) model of definite reference repair), CA analyses (initiated by Sacks) focus primarily on the *sequential* unfolding of troubles in understanding. They provide a detailed structural characterisation of these repair procedures.

Schegloff (1977) highlighted the way in which the repair initiation stage (the signalling of a problem) is distinct from the repair itself, and how both of these can be carried out either by the original speaker (self) of the trouble source turn (henceforth TST) or another person (other).

He pointed out the way speakers may foresee a potential, prospective trouble in the interlocutor's understanding, and repair the problem by restating or further qualification of part of their turn before the floor was ceded to the interlocutor, thus performing a self-initiated self-repair in the same turn.

If, on the other hand, the turn is released by a speaker as complete, the interlocutor now has a chance in the turn immediately following the TST, either to repair the problem herself by correcting some part of the TST (other-repair), or to initiate repair resulting in a meta-communicative insertion sequence ending by the turn containing the repair itself (other-initiated self-repair).

Jefferson (1983) further distinguishes within other-corrections in the next turn, between 'exposed' and 'embedded' corrections. The former occurs when the other-correction results in a break in the talk-in-progress, in the sense that in the ensuing talk the corrected participant, pro-

vides ‘accountings’ of why the original term was used. Embedded corrections, in contrast, occur when the correction is ‘adopted’ by the corrected participant, in a **sequentially relevant** turn following the other-correction, thus not breaking the talk-in-progress.

As regards other-initiated repairs or Clarification Requests (CR), see Purver (2004) for a full empirical corpus study of the different forms such CRs can get, and the range of semantic readings thereof. As indicated above in section 2.4.2, we should point out that repair initiations act effectively as the first pair part of an adjacency pair, where the second pair part is the repair itself.

Schegloff (1992) distinguished a further class of repairs; repair-after-next-turn. These occur when it is an interlocutor’s - proffered as - conditionally relevant response (the acceptance phase in the grounding cycle), which reveals to the speaker of the TST, that the interlocutor has misunderstood it. The speaker of the TST then has the opportunity to perform a self-repair in the next turn. These usually take the form of “no I didn’t mean A I meant B”.

The following schema summarises the different types of repair and their relative positioning with respect to the TST (Schema adapted from Levinson (1983); Schegloff (1992, 1997)):

- | | |
|-------|--|
| (2.5) | <ul style="list-style-type: none"> (1) A(TST): There are two opportunities for self-initiated self-repair: 1. Immediately after the error; 2. At the end of the turn. (2) B: Other-initiation or other-repair. (3) A: 1. Self repair by A upon other initiation in turn 2; 2. Third position self repair upon realisation by A of B’s misinterpretation of A’s turn 1, revealed in the response by B in turn 2. (4) B: Fourth position, other initiated repair; B realises after hearing A’s response in 3, that she had misunderstood A’s turn. |
|-------|--|

Although this thesis will not be dealing with repair phenomena specifically, it’s worth noting that most of the conversations or the segments therein on which these analyses of repair are based, involve only 2 participants as also evident in the scheme above. It is unclear how these repair mechanisms would scale up in multi-party dialogues.

We will return to this discussion, in section 3.9 of the next chapter, where we briefly discuss some of the questions that arise concerning repair phenomena in multi-party dialogues.

2.6 Coordination between, and Interrelatedness of the behaviour of interactants

It has long been known that when one interacts with others, one accommodates them by way of adapting to their needs and to the situation at hand, across a wide range of behaviours. One may act in a range of fashions or talk in different styles depending on the social situation/context in which one finds oneself. Indeed as Erving Goffman puts it, one's 'projected self' is potentially different in each case. But the same goes for those others that one is interacting with. In other words, they too, adapt their behaviour to ours, and hence are affected by the way we behave. So in interaction, the influence that we exert on one another is a reciprocal one.

In this section, we'll be concerned with some of the ways in which the behaviours of interactants are interrelated and how they influence one another locally, on a moment by moment basis. We will be looking at some of the existing literature relating to such coordination phenomena arising in social interactions, spanning both linguistic and non-verbal levels of behaviour. It will turn out that with few exceptions, all of the literature on behavioural coordination/alignment is based on the analysis of dyadic interactions.

2.6.1 Synchrony/Alignment in non-verbal behaviour

There is a relatively large body of empirical evidence to do with how one's immediate behaviour is directly affected by the behaviour of others present. Research in the social psychology literature, to do with the interrelatedness of behaviours of those involved in the same interaction, dates back to Condon and Ogston (1966).

The study in Condon and Ogston (1966), involved frame by frame microanalysis of recorded video and audio of dyadic interactions. The first problem in any such study is to determine the units of analysis according to which similarity in behaviour can be construed. To this end, Condon developed the notion of a *process unit*, which is roughly defined as a sequence of frames during which the directionality of the movement of one or more body parts is maintained. It is important to note here that if this involved more than one body part, then the inertial boundaries of the movements of each would need to coincide with those of the other body parts that moved, as though they were all dancing together to the same rhythm. Indeed, the rhythm here was being dictated by the speech that was simultaneously being produced. They also annotated the frames for the phonetic units that were being uttered during them. They found that the auditory boundaries coincided with those of the movements of the body parts. So far, this involves

only one person which is why the phenomenon is referred to as *Self-Synchrony*. *Interactional Synchrony* occurs, when the boundaries of the process units of the behaviour of the speaker, coincides with those of his interlocutor. Amazingly, it was found that in their data, interactional synchrony was the rule. Moreover, synchrony was maintained even when the participants were not paying visual attention to one another. Thus, it was all happening primarily via the auditory channel. They conclude by stating that interactional synchrony is an invariant characteristic of normal/non-pathological human interaction.

Condon's work was later followed up by Kendon (1970). Whereas Condon thought that synchrony was a pervasive, invariant property of normal human behaviour in interaction, Kendon's study (using the same methodology as Condon) shows that there are degrees or kinds of synchrony. This standpoint implies that synchrony may actually serve particular communicative functions in interaction. For instance the more 'strong' kind of synchrony found in Condon's study contributes, Kendon claims, to a general feeling of rapport between the interactants. Degree of synchrony is also said to correlate positively with degrees of engagement/attention in the interaction on the part of an interlocutor and hence could serve to regulate turn-taking. This leads Kendon to suggest that the task at hand is to describe the situations in which synchrony may or may not occur.

Specifically relevant to our concerns, is that Kendon's data in fact, comes from multi-party dialogues. We will not go into any more detail here, but in the later discussion of participation framework where this is more relevant. Suffice it to say that degrees of synchrony may serve to distinguish between the (normatively appropriate) behaviour of participants in different roles (for example primary vs. peripheral participants) in face to face interaction, which will be one of the main topics in our later discussion.

Also pertinent to non-verbal coordination, is Kendon's F-formation system (Kendon, 1976), which describes how people in face to face conversation manage to create and maintain the shared physical space between them. The F-formation as a 'ratificational device' (Goffman, 1981) will be more appropriate once the context is properly set in the next sections. We will not go into more detail here.

Condon's method of analysis (and Kendon's) has not remained uncriticised. In two papers, (McDowall, 1978b, 1978a), the first experimental tests of Condon's results were undertaken. It is concluded firstly, that the methodology of frame by frame analysis is not sufficiently replica-

ble, and secondly, that Condon's results are not more than a random coincidence of body part movements, meaning that the likelihood of these coincidences is not greater than what would be expected by chance. McDowell's claim is thus that interactional synchrony is not a genuine phenomenon. Gatewood and Rosenwein (1981) however, review of McDowell's work, concluding that McDowell's results are based chiefly on a misunderstanding of what Condon intended by the notion of a process unit, and hence that McDowell's experimental methodology is not a real test of the phenomenon of interactional synchrony. The details of this ongoing debate will not concern us any further in this thesis.

The kind of work alluded to above has more recently been taken up by other researchers. Synchrony has been defined overall as the matching of behaviours, the adoption of similar behavioural rhythms, the manifestation of simultaneous movement and the interrelatedness of individual behaviours (Bernieri & Rosenthal, 1991). Research has shown synchrony and non-conscious mimicry of behaviour to be related to positive affect (Bernieri, Reznick, & Rosenthal, 1988), interpersonal liking and smoothness of interactions (Chartrand & Bargh, 1999). This is often referred to in the literature as the coordination-rapport hypothesis. Most such studies found that degrees of affiliation with the interlocutor positively predict degrees of non-conscious mimicry. However, Chartrand and Bargh (1999) show that people mimic their interaction partners' mannerisms even when they were up to that point complete strangers and where it is unlikely that there is any pre-existing rapport or a goal to develop any future affinity. This has been explained by positing the perception-behaviour link hypothesis, whereby perceiving some form of behaviour in your interlocutor activates the same behaviours at an immediate/local level in you.¹ Despite all of this, Lakin and Chartrand (2005) among others, have shown mimicry to be stronger when a person has been socially excluded. So, the link between affiliation and mimicry/synchrony is still not completely straightforward in the social psychology literature.

2.6.2 Linguistic Coordination

So far in this section, we have only considered coordination/synchrony in non-verbal behaviour. A good starting point to move the discussion on to verbal/linguistic behaviour - dialogue - is Giles' Communication Accommodation Theory (Giles & Coupland, 1991). Analogously to the coordination-rapport hypothesis outlined above, according to the theory, interactants adapt to

¹We will shortly discuss a linguistic counterpart of this hypothesis, namely Garrod and Anderson's (1987) input/output coordination principle

each other's behaviours to create, maintain or decrease the social distance between themselves, in order to promote social approval or communicational efficiency. This can be done at various different levels, linguistic or otherwise, including e.g. code switching, modification of accent, idioms, dialect, gaze, pitch (e.g. see Giles and Smith (1979)), and word choice (Niederhoffer & Pennebaker, 2002).

As regards the coordination-rapport hypothesis alluded to in the previous section, Niederhoffer and Pennebaker (2002) find a rather surprising result. Although they had initially hypothesized that lexical matching (not defined as using the same words but words that belong to the same categories, e.g. articles, positive, negative, sex related etc.) would be positively correlated with 'clicking' or rapport, they found that while there was always a high correlation on both turn-by-turn and conversational levels between the word categories used by the two participants, there was no correlation between degree of lexical matching and rapport (judged both by outsiders and the participants themselves). This leads them to suggest a rather different hypothesis, namely coordination-engagement. It would help to think for example of a couple who are having a serious row. Intuitively their non-verbal behaviour would be quite synchronised, while they would also match each other constantly in the use of negative words. They are highly engaged in this interaction although something the opposite of rapport holds at that point. These results, Niederhoffer suggests, are not inconsistent with Communication Accommodation Theory, as the theory allows for interactional complexity. This means that in the model, interactants can converge on some aspect of communication while diverging on another, to manage their identities. After all, different people do come across as different and this seems to be, at least partly, intentional.

Particular to our concerns here is the success of verbal communication as a consequence of coordination/synchrony/alignment in dialogue, which brings us to the *interactive alignment model* (Pickering & Garrod, 2004). The central idea in this psychological model, is *representational alignment*, whereby at various levels, the representation used by a speaker *primes* similar representations, in subsequent utterances by the interlocutor. If A uses some representation, then his addressee is very likely to use the same representation in his response to A. This priming takes place at all levels of representation, including phonological (Goldinger, 1998), lexical (Niederhoffer & Pennebaker, 2002), syntactic (Brannigan, Pickering, & Cleland, 2000) and semantic. On this account production and comprehension are not psychologically independent mechanisms (as traditional generative linguistics assumes they are). Garrod and Anderson

(1987) describe this via the input/output coordination principle whereby an interlocutor aligns the representations he uses in producing an utterance, with those he has just used in interpreting/comprehending the previous utterance addressed to them.

It is generally alignment at the level of the situation model (top-level semantic representation) that leads to successful communication. Pickering and Garrod claim that alignment at lower linguistic levels percolates and leads to alignment at higher levels, and that priming is the central automatic, resource-free mechanism, responsible for such alignment, and successful mutual understanding.

It should be noted that the most important difference between this model and Clark's grounding model, is that whereas successful communication in the interactive alignment model, is a consequence of automatic, resource-free priming mechanisms, in the grounding model, successful communication would require computationally expensive processes of reasoning about another's beliefs and intentions; 'other-modelling', and hence, Pickering and Garrod argue that the latter cannot be a good characterisation of spontaneous dialogue. They claim that inferencing about another's mental state (beliefs and intentions) is a participant's last resort, that is, it is only invoked when everything else fails to secure mutual understanding. This criticism is one that has recently been increasingly directed at Grice's intentionalist view on communication, and Clark's adaptation of it in the grounding model (recall also Keysar's criticisms sketched in section 2.3.7 above).

Although alignment appears to be a pervasive characteristic of successful, effective interactions, the claim that alignment at one level leads to alignment at others seems to be too strong. We have already seen that it is at odds with Neiderhoffer and Pennebaker's (2002) conclusions with regards to convergence on one level while not precluding divergence on another, hence enabling the management of identities. Garrod and Clark (1993) also found that while younger children were aligning on the words/expressions used, the expressions did not have the same reference. So lexical alignment here did not lead to semantic alignment, on the contrary, it was detrimental to successful communication. This leads Garrod and Clark to suggest that the automatic compulsion to align had to be actively suppressed in such cases.

It's perhaps useful to recall the tension between Kendon and Condon's approach, summarised in section 2.6.1. Namely that Kendon believes that synchrony, rather than being characteristic of all non-pathological interaction, serves particular purposes such as the regulation of turn-taking,

or to signal affiliation. Linguistic alignment too, can serve multiple purposes. For example lexical alignment or repetition can serve to accomplish embedded repairs, initiate repair, signal affiliation and rapport etc.

Intuitively it appears that if it were the case that we aligned at all levels in what we say to one another, it would mean that when having a conversation we would be stuck in a loop. Don't we switch perspectives, paraphrase another's utterances to give them the evidence of understanding they require? If someone repeated verbatim what I say to them, then I'd have the least degree of certainty about whether they've understood me. Aphasics can repeat/shadow even when production and comprehension are heavily impaired (Whitaker, 1982). Conversely, if one paraphrases what I say involving a radical change in the situation model used, wouldn't this serve as a strong basis for the assumption that they've understood me? It appears as though we may exploit representational misalignments - all those repairs and edits that we do of one another's utterances/meanings - as a means to situational alignment and hence successful communication.

It's crucial here to note that most of the literature alluded to above - except Kendon's work which will be sketched in the next chapter - uses data from social encounters in which there are only two participants. We will later discuss how linguistic priming might also be at work in multi-party dialogues and how it relates to participatory status.

2.7 Computational/formal accounts of dialogue context

We will now, quite summarily, consider two of the most prominent computational models of dialogue processing/context, namely Ginzburg's KoS framework (Ginzburg, 1996) later extended by Larsson (2002) and others, and Traum's (1994) computational model of grounding. Traum's model is a direct formalisation of Clark's grounding model, while Ginzburg's KoS, uses various insights from both the work of Clark and colleagues and the CA tradition on the integration of meta-communicative interaction (repair) into the model.

As formal models of dialogue context, they need to satisfy the following requirements:

1. They should capture the highly constrained set of next moves that are coherent/felicitous at any given point in conversation, viz. they specify a protocol for interaction, in terms of preconditions on the state of the corresponding context model, that should hold if a certain dialogue act or utterance type (e.g. short answer, acceptance etc.) is to be 'felicitous' at that point. Moreover, for each utterance type they specify the change/update that the utterance

brings about in context. So we might say that such models are intricate state transition systems, where utterances are the triggering events, or equivalently the edges between the states in the transition diagram.

2. They should facilitate the resolution of context-dependant utterances: VP-ellipsis, Clarification Ellipsis, Short Answers and the like.

2.7.1 A very brief sketch of Ginzburg's KoS

As mentioned above in passing, accounts of contextual change can be traced back to the 'Classical' view of Stalnaker (1978). On this view, context is defined at time t as the set of assumptions that the dialogue participants hold *commonly* at t (common ground).

How would an utterance modify the context as defined by Stalnaker? He suggests that the propositional content of an assertion made by a participant is added to the context provided that his interlocutors accept it. In other words a participant can either accept or reject an assertion. In the latter case the context will remain unchanged.

However, Ginzburg (1996) argues that such a view would not provide a realistic model of contextual updates, since it doesn't properly characterise the highly constrained *discursive potential* of dialogue after an utterance has been produced. More specifically Ginzburg rejects Stalnaker's simplistic Accept/Reject dichotomy.

Stalnaker's simplistic definition of context is essentially an unordered set of propositions. If the relevance or felicity of a next utterance were to be characterised in terms of this set alone, each element in the set would play an equal part, being of equal importance. The relevance of the next turn would not depend, in this model, on the order in which the propositions were added to the set. An important property of dialogue however, is, as sketched in section 2.4, sequential coherence. This corresponds to the Conversational Analytic notion of *conditional relevance*. The last utterance greatly constrains the set of relevant follow-ups to that utterance. Moreover, the semantic content of elliptical follow-ups (Short Answers, VP-Ellipsis, Acceptances, Rejections etc.) are often resolved against the context of the last utterance, i.e. the semantic content of an elliptical answer to a question can only be recovered with reference to that question. Hence Ginzburg suggests that the *Latest-Move* be retained as a structural feature of a participant's representation of context. The reaction to an utterance could also be an attempt to initiate repair, or more broadly to open an insertion sequence (the CA notion outlined in section 2.4.2). But the

opening utterance of the insertion sequence itself is again restricted by and interpreted against the context of the Latest-Move. The Latest-Move is therefore also needed, to characterise the range of repair initiations (or Clarification Requests) that can follow.

It is worth noting that in order to account for certain syntactic and phonological parallelism phenomena, it is not just the semantic content of the Latest-Move which needs to be projected into context, but also the syntactic and phonological information associated with it. This is in fact one of the motivations behind using HPSG (Head-driven Phrase Structure Grammar), a constraint-based grammar, as the formal framework in which the theory is couched. Within this framework the phonological, syntactic and semantic information associated with an utterance are represented, *in parallel*, meaning that constraints on production and parsing could be expressed using a combination of all such information. We will not go into any more detail but here's a simple example of syntactic parallelism between the Latest-Move and a Clarification Request follow-up:

- (2.6) | (1) A: I saw her at the party.
 | (2) B: her? ["she?"] would not acceptable here as "her" is the object of
 | the verb in the previous utterance]
 | (3) A: I mean Mary.

Now recall that an insertion sequence is preliminary to the discussion of the question which prompted it. Hence before that original question can be addressed, the issue(s) raised in the insertion sequence (possibly a repair) need to be resolved. This is why *QUD*, a partially ordered set (often a stack) of questions, is introduced as an additional contextual feature. Before a question can be answered, its more local sub-question (a question on which the original one depends) needs to be addressed, and popped off the stack. The currently most salient question, the top of *QUD*, which any next utterance should be 'about', is dubbed *QUD-Maximal*.

To summarise, the framework structures the context into the following features:

- **FACTS:** The set of propositional facts accepted commonly by the participants, more formal but analogous to Clark and Schaefer's notion of common ground.
- **Latest-Move:** An attribute representing the linguistic sign (in HPSG) associated with the latest move made.
- **QUD:** The feature which keeps track of all the currently discussable questions. *QUD* is a

partially ordered set, the maximal element of which (dubbed QUD-maximal) is what the resulting protocol allows the next utterance to be ‘about’. The next utterance can also be a question on which QUD-maximal ‘depends’.

The above triple is referred to as the Dialogue Game Board (DGB), which is effectively a formal view of the participants’ common ground. Each dialogue participant’s view of the common ground is thus a record containing the attributes outlined above.

Depending on the values of these attributes at a moment in time, the DGB has a configuration. One could think of the model as an intricate state transition system. Patterns of conversational interaction can in this manner be described in terms of the evolution of the DGB. Now it is *Conversational Rules* (Larsson, 2002), that determine the mappings from one DGB state/configuration onto another.

Here are the set of Conversation Rules (stated informally, instead of in HPSG notation), specifying the (legal) mappings from one DGB state onto the other governing dyadic interactions:

- **QUD-Specificity (QSPEC):** given $\text{MaxQUD} = q$, one can make an utterance which is q -specific.
- **AskQUD Update:** given $\text{LatestMove} = \text{Ask}(A,B,q)$, q becomes QUD maximal
- **Assert QUD Update:** given $\text{Latest-Move} = \text{Assert}(A,B,p)$, $p?$ becomes QUD maximal
- **Accept:** given $\text{LatestMove} = \text{Assert}(A,B,p)$, B can make utterance such that $\text{LatestMove} = \text{Accept}(B,A,p)$.
- **UpdateFacts + DowndateQUD:** Given $\text{LatestMove} = \text{Accept}(B,p)$, conjoin p with FACTS, downdate $p?$ and all other qs from QUD resolved by FACTS

What constitutes a q -specific utterance is explicated by Ginzburg using the notions of question ‘dependence/influence’ and proposition ‘aboutness’, intricate semantic notions that we will not discuss here. Suffice it to say that, informally, an utterance is q -specific either if it’s a partial answer to q or a sub-question of q . The important point is that the notion of question-specificity is intended to provide semantic constraints on the set of next ‘coherent’/felicitous moves, while facilitating ellipsis resolution.

Recall now what we mentioned in passing regarding interactional protocols, in section 2.1.2. We mentioned that interaction protocols, in Ginzburg’s model, arise as a result of the composition

of Conversational Rules such as the above. These can be composed if the set of preconditions of the second rule is a subset of the set of effects of the first. The above set of rules, for example, give rise to the following interaction protocol for Querying:

LatestMove = Ask(A,q): A has asked B the question q

A:

push q onto QUD;

release turn;

B:

push q onto QUD;

take turn;

make max-QUD-specific utterance;

release turn.

Such protocols then (there are others for Assertions and Greetings) are intended to provide constraints on what constitutes a coherent dialogue. Interesting questions arise here: what if these constraints are violated? How do we characterise the vastly context-dependent and variable nature of the implicatures that would be generated in such circumstances? And what about the effects of such violations on the ‘rules’ themselves? On the notion of ‘specificity’ to a some question which has received an ‘unlicensed’ answer? Although these questions are notoriously difficult to address at an adequately formal level, they do point to the need for a model in which such norms or rules are emergent rather than, in some sense already there.

Relevant to our concerns in this thesis is that this model, and the interaction protocols induced thereby, arrange possible actions of only two interactants; it is dyadic. In the next chapter, once we’ve introduced the empirical differences between dyadic and multi-party dialogues, we will return to our discussion of this model, and the attempts which have been made at some transformations to handle multi-party interactions. We will then raise some questions as to whether the empirical adequacy criteria, which have given rise to the proposed transformations, bear closer inspection.

2.7.2 David Traum: Grounding and Collaborative Conversational Acts

Traum (1994) presents a computational theory of grounding, where some problems with the grounding model are brought out and dealt with. Although the strength of evidence principle is invoked to rectify the problem of infinite recursions in accepting acceptances, Traum (1994)

claims that the model is still unclear on whether contributions are ever complete. Indeed Clark and Shaefer do suggest that some acceptances need not themselves be accepted, and hence aren't presentations. However, this is really as far as they go in specifying the conditions under which the latter is the case.

Also, Traum claims that it is often unclear whether an utterance belongs to the presentation or the acceptance phase: "we often need to look at large segments of the conversation, both before and after an utterance before deciding how the utterance fits in" (Traum, 1994). Note that this is not an empirical problem but a computational one, as people do seem to know in practice how a particular utterance fits in.

One way or the other, Traum solves such problems, by replacing the presentation and acceptance phases with more atomic grounding acts (a single stratum in Traum's (1992) formulation of Conversational Acts introduced below, comprised of the actions which contribute to 'groundedness' in general) and on that basis develops a finite state automaton which will at any stage in the dialogue, reflect whether a Discourse Unit (the term replacing Clark and Shaefer's notion of a complete contribution, also the level of structure at which conversational content is grounded so that we always speak of the groundedness of a discourse unit) is already grounded and if not, what it takes to ground it. I will not go into the details of how exactly this is done, due to scope limitations. Suffice it to say that the problems with the presentation model brought out by Traum are really only significant if we are to develop a computational protocol for grounding in terms of exactly what actions are required by which participants (agents) to reach a certain desired state of the grounding cycle. Clark and Shaefer's account of grounding seems empirically adequate, so far as we consider only dyadic dialogue.

Collaborative Conversational Acts

The theory of Conversational Acts (Traum & Hinkelman, 1992), is a generalisation of Speech Act theory, in dialogue. Conversational Acts are organised into a hierarchy, such that each stratum contains actions relevant to a single discourse level, which correspond to larger and larger chunks of discourse as we go down the hierarchy. Traditional speech acts are retained in the third of these strata, corresponding to actions taken at the level of the Discourse Unit. The crucial consequence of such formulation is that an utterance containing a speech act, does not take effect until its corresponding Discourse Unit has been grounded. In this manner, discourse is viewed as

a collection of joint speaker-hearer actions, unlike traditional speech act theory which assumes autonomy on the part of the speaker and hearer. We should see that Conversational Acts are essentially the result of incorporating grounding theory into the theory of speech acts. If A asks B to pass the salt, the request does not take effect until they come to the mutual understanding that the utterance containing the request has been understood.

The collaborative approach to the notion of action in general has not remained uncriticised. Below is one such criticism.

A critique of collaborative action

A critique of Clark and Shaefer and also particularly of the consequent collaborative theory of conversational acts in Traum and Hinkelman (1992), can be found in Allwood (1995). This paper criticises the central contribution of Clark and Shaefer's grounding model, namely that speech acts in conversation are **always** performed collaboratively (vs. unilaterally). In Traum's theory, ordinary/traditional speech acts (high-level stratum in Traum's taxonomy of conversational acts) do not take effect in dialogue until they have been grounded, i.e. both agents need to take action.

According to Clark and Schafer it is not possible to perform a communicative act, without an appropriate reaction on the part of the interlocutor. Allwood's (1995) argument against this, is based on cases where dialogue turns are not yielded. For example, when a querier answers her own question and no one else addresses that question, something which happens not infrequently and can be done quite smoothly. Or, "I was referring to Bertrand Russell but she did not hear me". The point is that one has done something in one turn alone, regardless of whether or not the interlocutor responds in any kind of fashion. 'At least as far as the conception of action which surfaces in ordinary language is concerned, communicative acts need neither necessarily be resultative [have an effect] nor intentional and the fact that they need not be resultative, in turn, means that they need not be collaborative' (Allwood, 1995). In summary, the criticism is that theory of Collaborative Conversational Acts runs counter to our usual intuitions about action.

Chapter 3

Scaling up: Multi-party conversation

3.1 Introduction

In the previous chapter, various aspects of two-person dialogues were sketched, and many discussions deferred until now. In this chapter, we'll summarise some of the existing literature on multi-party dialogues, with the aim of bringing into focus the difficulties we face in scaling up theories of dyadic dialogue to multi-party dialogue. Some of the discussions in the last chapter will be picked up again here, in the context of multi-party dialogues. The specific questions that this thesis is concerned with will be highlighted.

The central notion that we will be concerned with in this chapter is that of the *Participation Framework* in multi-party dialogues, constituted by roles other than Speaker and Hearer, an issue which does not really arise in studies of dyadic dialogue. In particular, we will be concerned with the grounds on which such roles are distinguished. Everything else will count towards the characterisation of these roles and the difficulties therein, in terms of patterns of turn-taking, grounding, conversational responsibilities, sequential coherence and coordination mechanisms in general.

3.2 Participation Framework: Critiques of the dyadic model of communication

There has always been a bias in the literature (linguistics, sociology, semiotics etc.) towards taking dyadic dialogue as the “canonical” form of communication. Levinson (1988) even speaks in this regard, of a ‘straitjacket’. He argues that the success of the research on dyadic interac-

tion has been at the price of limiting the types of social situations being considered. Multi-party situations, particularly in certain societies are in fact a lot more common than dyadic ones. Presumably, the assumption that has been made by researchers is that all forms of communication can ultimately be somehow reduced to the dyadic Speaker/Hearer model. But:

All such schemes appear to agree in either taking the standpoint of an individual speaker or in postulating a dyad, speaker/hearer (or source/destination, sender/receiver, addressee/addressor). Even if such a scheme is intended to be a model, for descriptive work it cannot be. (Hymes, 1972)

Levinson (1988), however, points out that ‘beyond one or two suggestions, Hymes does not offer us a better scheme’. Goffman (1981) is the first researcher to suggest the need for decomposing the notions of Speaker and Hearer into their smaller and analytically coherent constituent concepts. Decomposing the Speaker category yields what Goffman calls the *production format* which designates possible kinds of speakers, and Hearer into what he refers to as the *participation framework*, designating possible kinds of recipients. Rightly, according to Levinson, these terms are misleading. We will therefore use Levinson’s alternatives throughout. These are respectively, *Production Roles* and *Reception Roles*, reserving the term Participation Framework for the union of the two role sets, instantiated with respect to what Levinson calls an *utterance-event*. This latter term will be motivated and defined once we’ve introduced Goffman’s analysis.

3.2.1 Footing

Before we begin considering Goffman’s and subsequently Levinson’s proposals, we need to introduce the notion of *footing*. Goffman himself wasn’t completely clear on what ‘footing’ exactly meant. But vaguely, issues of footing arise when “participant’s alignment, or set, or stance, or projected self is somehow at issue” (Goffman, 1981). As alluded to above with regards to behavioural coordination, the projection of such a stance can be manifested through various means including posture, gaze, and more importantly here, linguistically, including via pitch, accent, dialect and word choice. Changes in footing can be manifested through changes in any of these aspects of behaviour. For example, Blom and Gumperz (1972) note that in the Norwegian community, there are effectively separate codes, corresponding roughly to formal/official and colloquial manners of speaking. They observe that Norwegian people often switch between these,

depending on the social context/situation in which they find themselves. For instance, they describe a situation where an outsider steps up to a group involved in close conversation, at which point there is suddenly a shift from code to code and also in the posture of the group. Another example from Gumperz, shows that speakers switch codes in conversation when moving on to talk about serious business from intimate greetings and exchanges about their families. This occurs in the same conversation between the same two people.

It is perhaps a surprise then that Goffman speaks often as if footing always reduces to issues of participation status: “the alignment of an individual to a particular utterance whether involving a production format, as in the case of a speaker, or solely a participation status, as in the case of a hearer, can be referred to as his *footing*” (Goffman, 1981). It seems that participation framework is for Goffman, a structural analysis of certain special changes in footing. Once we’ve introduced fully the notion of participation framework (which is really what we’re concerned with rather than the more general notion of footing), whereby the speaker/hearer categories are generated, it will appear that not all changes in footing can be accounted for in terms of changes in the participation framework. To see this, recall the latter example above from Gumperz in the previous paragraph, in which there is a change in footing as a result of a change in the kind of talk being had. Indeed it seems that footing isn’t an issue arising exclusively in multi-party dialogues, but one arising in any social interaction including dyadic talk.

3.2.2 Production Roles: Decomposing the *Speaker*

Being a speaker in the ordinary sense (which is the most pervasive) is, as Goffman (1981) suggests, really to be in three roles simultaneously:

1. *animator*: The utterer of the message.
2. *author*: The composer of the message.
3. *principal*: The ultimate source of the message.

It is important to recognise that these roles do not necessarily coincide. For example a spokesman is the animator and author of his message but not the principal. Or in case someone directly quotes another, he is the animator but neither the author nor the principal.

Though the production roles as formulated above, are not without their own problems, we will not in this review be much concerned about the distinctions made between different kinds

of speakers, as this doesn't really bear on the issues arising in theories of multi-party dialogue. We will therefore not expand Levinson's revision of Goffman's production format. This is also because Goffman's categories (above) seem empirically sufficient (though not as systematically formulated as Levinson's).

3.2.3 Reception Roles: Decomposing the Hearer

Here's the scheme proposed by Goffman for different kinds of hearers:

- *Ratified*

1. *addressed recipient* 'the one to whom the speaker addresses his attention and to whom, incidentally, he expects to turn over his speaking role'.
2. *unaddressed recipient* the rest of the 'official hearers', who may or may not be listening.

- *Unratified*

1. *bystanders* 'inadvertent', 'non-official' listeners.
2. *eavesdroppers* 'engineered', 'non-official' followers of talk.

However, "although Goffman's categories are a notable advance on earlier schemes, they do not seem sufficient" (Levinson, 1988). Consider the following excerpt from Sacks, Schegloff, and Jefferson (1974), adapted from Levinson (1988):

- | | |
|-------|---|
| (3.1) | <ol style="list-style-type: none"> (1) Sharon: You didn' come tuh talk tuh Karen? (2) Mark: No, Karen- Karen' I're having a fight, (0.4) after she wnet out with Keith an' not with (me) (3) Ruthie: Hah hah hah hah (4) Karen: Wul, Mark, you never asked me out |
|-------|---|

According to Goffman's reception roles, with respect to Mark's utterance, Karen is designated as a *ratified unaddressed recipient*. However, she is clearly more than that since we'd expected her to respond to Mark's utterance as she does. She is here the intended recipient of that utterance. This is a case where the addressee (Sharon) does not coincide with the intended

recipient of the message. What would Karen's role be if she were unratified and hence could not respond as she does?

To increase the empirical coverage of Goffman's scheme Levinson (1988) recognises 4 different features in hearership. Depending on the values of these features which are true or false of a non-speaking party with respect to an utterance-event (defined below), they will generate a set of possible reception roles. The features are as follows:

- *Address*: Is the utterance addressed at the party?
- *Recipient*: Is the participant the ultimate target of the utterance?
- *Participant*: Is the participant ratified?
- *Channel-Link*: Is the participant in the perceptual range of the utterance?

For example, in Karen's case above the Recipient, Participant and Channel-link features would be turned on, but not Address, thereby generating the category *indirect target*. If she were unratified in Goffman's sense, she would also have the Participant feature turned off, thereby generating the *targeted overhearer* category. While generating all of Goffman's original categories, this scheme seems to solve the problems faced by Goffman's in cases like the example just discussed or chaired committee meetings.

But what empirical evidence is there that such a scheme is adequate? More importantly, how can we be sure that participants really make these distinctions and that they are the correct ones? As Goffman points out, "linguistics provides us with the cues and markers through which such footings are made manifest", and these cues and markers are exactly what Levinson (1988) aims to reveal in different languages. He justifies these categories by linking the linguistic devices employed in different languages with the distinctions possible both at the production and the reception end. However, we need to see that Linguistics does not exhaust these cues and markers:

Some of these are purely non-verbal such as those that Kendon (1976) discusses regarding patterns of how people orient themselves spatially towards one another in order to manage ratification, in face to face conversation. We will discuss this further below when we consider the problematics of ratification.

Moreover, as mentioned previously, Kendon (1970) shows that different degrees or kinds of synchronous behaviour serve particular communicative functions. Participants display synchrony

with the speaker to show that they are attending to him. Different degrees of synchrony may then serve to distinguish participants in different roles (e.g. primary vs. peripheral). Additionally, what Kendon refers to as ‘mirror synchrony’ occurs when a non-addressed recipient wishes to draw the attention of the speaker to facilitate speaker switch, or in the present context, we could interpret this directly as the management of changes in the state of the participation framework. Moreover, gross changes in posture occur on the macro level among the non-addressed recipients, coincident with points of speaker change that do not involve any change in the interactive axis (when the two participants talking to each other remain the same). We should see however, that as Kendon’s data comes from focused encounters, it is not completely clear how, for example, overhearers would be distinguished in this sense. But intuitively, since synchrony is a non-conscious phenomenon, an overhearer would synchronise with the speaker if she is paying attention to them.

Now that we’ve presented Goffman’s and Levinson’s proposals for participation structures, we need to consider the frame of reference, the unit, according to which these roles are to be determined. As alluded to above, this is the *utterance-event*.

3.2.4 Frames of reference

There is, in general, a systematic ambiguity in the use of terms like speaker, hearer, audience etc. This is the speech-event/speech-act ambiguity, in the sense that all participant roles can be identified with respect to a speech-event as well as a speech-act. It is easy to see that the two senses do not always coincide. For example, a lecturer is the speaker in the speech-event sense, even when a student is asking a question. This is in fact a distinction that, according to Levinson (1988), Goffman fails to consistently make (though he is well aware of it). Levinson points out in this regard, that ‘we are specifically concerned with the *utterance-event* use of terms like speaker and addressee, for the speech-event usage is parasitic on this primary usage’. In other words, what we need to be focusing on, in describing interaction, is a *bit by bit* analysis of participation through the course of a conversation: ‘the investigator is wise to drop the idea of even attempting to describe a global interaction, and instead be content with taking a more modest approach and examining only certain moments of talk’. (Kerbrat-Orecchioni, 2004)

Before we move on to consider how Levinson defines an utterance-event, let’s look at why it is generally so difficult to talk about conversation as a unit of interaction, and how definitions of conversation as such, run into fundamental difficulties when we come to consider multi-party

gatherings.

3.2.5 Problems of Multi-party Conversation as a unit

Theories of human interaction use conversation as a basic building block. However, as Kendon (1976) points out, this is not an unproblematic notion:

A central methodological problem in the study of the behaviour of face-to-face interaction is that of defining the structural units in terms of which it is organised. In particular, it is of great importance to be able to **delineate**, in terms of the organisation of observable behaviour, distinct units of interaction which can then be analysed into their components.

The notion of conversation as a unit is supposed to group together a set of people and communicative signals, verbal and non-verbal, which are considered together in an analysis. In dyadic dialogue this choice is relatively unproblematic. In multi-party dialogues on the other hand the boundaries of interaction become more difficult to both define and track. And this is essential if we are to understand the pragmatic structure of an unfolding multi-party interaction.

This methodological difficulty has lead researchers to concentrate mainly on what Goffman (1963) has called *focussed interaction*: “the kind of interaction that occurs when people gather close together and openly cooperate to sustain a joint focus of attention”. Goodwin (1981) also equates conversation with focussed interaction in this sense: “Despite the broad scope of the term, when used in this sense [in a loose way as an equivalent of talk or spoken interaction], conversation is but a special case of what Goffman (1963) has called focused interaction”.

Let us now explore the problems that this methodological bias in the characterisations of conversation as a unit lead to. Goffman (1981) defines conversation as:

a substantive, naturally bounded stretch of interaction comprising all that **relevantly** goes on from the moment two (or more) individuals open such dealings between themselves and continuing until they finally close this activity out. The opening will typically be marked by the participants turning from their several disjointed orientations, moving together and bodily addressing one another; the closing by their departing in some physical way from the prior immediacy of copresence. Typically, ritual brackets will also be found, such as greetings and farewells, these establishing

and terminating open, official, joint engagement, that is, **ratified participation**. In summary, a social encounter.

As can be seen in the above definition, it is the participants' physical orientation, movement and the use of ritual brackets that mark the boundaries of a conversation, and determine the persons who are officially participants in it.

Analogously, as alluded to previously, Kendon (1976) identified more subtle ways in which people use body orientation and movement to maintain an 'interactional unit' or F-formation: "the F-formation system serves as an important means of maintaining the separate identity and integrity of an interactional situation" (Kendon, 1976).

Nevertheless, as Goffman himself is aware, definitions of conversation as focussed interaction, as above, are contradicted in discontinuous strands of interaction where there is an 'open state of talk' (Goffman, 1981), i.e. when conversational contexts are 'left open' - questions left unresolved - and taken up later on, in the sense that what you say later, counts as 'pragmatically next' or a second pair part, to what was said before the conversation was 'broken'. A simple example is that of multi-focus gatherings, where one person may be a participant in more than one parallel conversation, as in the following constructed strand of talk:

- | | |
|-------|---|
| (3.2) | (1) Siavash: so where did you end up going last night? |
| | (2) Raha: Arash, where did you put my wine glass? [shouting from the kitchen] |
| | (3) Arash: It's in there somewhere. |
| | (4) Raha: oh ok it's here. thanks. |
| | (5) Arash: sorry yeah erm we went to the cinema. |

Importantly, Goffman's ideas about the participation framework, elaborated previously, also run into difficulty in such circumstances. A single participant can have more than one participatory status depending on the conversation relative to which her participant status is to be determined.

Perhaps the most difficult as well as the most general question in this regard is: What are the criteria by which an utterance can be regarded as a coherent extension of a given conversation? What constitutes **one** conversation as opposed to many? It seems that the pragmatic cohesiveness, or the dependency of the syntax/semantics of turns on prior ones, showing that some conversational context is shared by the participants, is a good candidate criterion.

In such terms, it is notable that in discontinuous, multi-focus strands of interaction, ellipsis or different surface forms of context-dependency can cross conversational boundaries. Arash's last utterance above has its antecedent in the first turn. For another example, consider the following (this is a real dialogue):

[I have had a prior conversation with Stuart regarding the BNC CD, which I promised to bring in. I arrive at the office . . .]:

- | | |
|-------|---|
| (3.3) | (1) Stuart: So? [This is resolved as "So did you bring the BNC CD?"]
(2) Arash: Yes.
(3) Stuart: [smiles]
(4) Arash: [smiles and gives him the CD] |
|-------|---|

In the above dialogue, Stuart could also have said: "So, did you bring **it**?", or "So, did you bring **the CD**?".

But what are the conditions which make all this possible? Certainly whatever it is that the utterance is referring to should be salient in the participants' shared-context. But what is the correct measure of salience? Distance? Clearly, the notion of distance cannot *just* be number of turns or time elapsed, since I could have had the prior conversation with Stuart 1 week before and this could still be possible. Another question is what if Stuart and I had started to talk about something different first (different topic)? Wouldn't Stuart have to avoid at least the highly elliptical "So?", since I would take it as a continuation of our 'current' conversation (the 'local' interpretation being preferred)? How do we investigate this in a systematic fashion?

In section 3.10 below, we will see how even within one and the same focused, multi-party interaction, the pragmatic structure of turns is different and more complex than in dyadic dialogues. The questions raised here and later in section 3.10 are central in this thesis and will be taken up in chapters 6 and 7.

3.2.6 Utterance-event

Participation Framework (PF) is the union of the production and reception role sets, instantiated with respect to an *utterance-event*. Levinson (1988) defines an *utterance-event* as "any stretch of a turn at talk during which there is a constant set of participant roles mapped into the same set of individuals". An utterance-event as such, is generally neither a turn at talk, nor a sentence.

To see this, consider Goodwin (1979)'s analysis of the sentence: "I gave up smoking cigarettes

one week ago today actually” in a recorded dinner conversation. This sentence is addressed at three different participants, as the speaker’s gaze moves from one guest to his wife to another guest, seeking a direct addressee. This order is not arbitrary. Unlike the guests, his wife already knows that he’s given up smoking but does not know that exactly one week has passed since. But “one should not tell one’s co-participants what one takes it they already know” ((Sacks, 1973); see also Grice (1975)), hence addressing only the second part to his wife. Because the state of the participation framework (the current negotiated set of roles) at any moment has an impact on the behaviour, linguistically or otherwise, of the speaker, Goodwin is justified to say that the sentence/turn is comprised of three distinct sections, which correspond to Levinson’s utterance-events.

3.2.7 The overarching pragmatic problem

Ideally, from a theoretical viewpoint, we would like to be able to fully instantiate the Participation Framework for any given utterance-event, so that for an utterance event UE we have $PF(UE)$ which assigns to each role a unique participant of the utterance-event. In the next section we will consider the difficulties of this task.

Let’s suppose, for the moment, that we manage to assign roles uniquely in an utterance-event. Our final goal is from the point of view of dialogue pragmatics, to incorporate the participation framework as a contextual parameter which will in turn determine the pragmatic ‘force’ of the utterance-event relative to the role occupied by the hearers thereof. In short we want to know how an utterance-event changes the context locally for each participant. So, given A ’s participant role with respect to an utterance-event UE , we need to characterise the set of obligatory, as well as optional next moves or reactions, by A , as sequentially implicated by UE . This of course will only be possible if we can assign interpretation of UE for A , relative to the existing shared-context between A and the speaker.

This is a very difficult task. The rest of this chapter will be dedicated to bringing into focus some the numerous sub-problems of this problem.

3.2.8 Issues in assigning participant role given an utterance-event

As outlined above, given an utterance-event we need to assign to each participant present in the conversation a participant role. we therefore need to know which of the features summarised in section 3.2.3 will have a positive value, given the utterance-event. We will consider each of them

in turn below.

address

We can determine the value of the address feature by taking into account visual address markers such as gaze, head movements and body orientation. Indeed Goffman defines addressee as ‘the person to whom the speaker directs his visual attention’. Non-visual cues include use of formal linguistic features such as second-person pronouns, or names. However, such address markers are not always present. Participants often recognise that they are addressees by mere reference to the semantic content of the utterance in question. For example it may be that the utterance includes information specific to a particular participant.

Recipient

The recipient feature on the other hand, is even more difficult to deal with since unlike the address feature, it cannot be characterised without reference to the semantic content of the utterance-event.

Ratification

When is a participant ratified? What are the procedures via which participants get ratified? Can we recognise different degrees of ratification? All this of course, comes down to the question of how one should interpret the ‘participant’ feature (Levinson, 1988).

Goffman defines ratified participants as those who are ‘officially’ members of the conversational group. This can be seen by how participants orient themselves towards one another for instance in terms of eye-contact, proxemics and spacial orientation (see e.g. Kendon (1976); Goodwin (1981)). However, there doesn’t seem to be complete consensus among scholars as to what exactly constitutes ratification. For example, Drew (1992) considers the jurors in a court room as unratified effectively because they are unable to speak or comment. Indeed this is quite strange since institutionally they are the intended recipients. Heritage (1985) says the same thing but regarding this time, the audiences of news interviews. But who is more ratified in a court room, jurors or spectators? Clearly jurors are in a sense more ratified since they are intended recipients of the utterances.

At the speech-event level, ratificational devices are perhaps as diverse as the speech-events themselves (Levinson, 1988). At the utterance-event level, speakers can, for instance, ratify a former overhearer by addressing her with an utterance relevant to the current conversation. Conversely, attempts by an overhearer to seek ratification may sometimes lead to explicit negotiation,

e.g. “Would you please stay out of this?”.

An example of a non-verbal ratificational device is Kendon (1976)’s F-formation: “An F-formation arises whenever two or more people sustain a spatial and orientational relationship in which the space between them is one to which they have direct and exclusive access”. An F-formation is said to serve as a means to maintain the identity and integrity of an interactional situation. Changes in the F-formation system occur for example when somebody joins or leaves the group. Pertinent here in particular, is when a person joins the group. It may be the case that the current members of the group feel that the ‘integrity’ of the situation will not be harmed if he joins, and hence are willing to ratify the new person. In this situation say if they were arranged in a circle, they will move aside a bit and let him in, while maintaining their respective orientations to the shared space between them, thus allowing the newcomer to orient to it in the same way, with the same degree of access.

Furthermore, at the utterance-event level, not unlike the ‘recipient’ feature, it seems that ratification too cannot be fully characterised without reference to the semantic content of utterances. Consider a not very uncommon situation in which there is a normal conversation between a husband, wife and a very close friend of theirs. The conversation develops into a row between the couple, at which point the friend recognises that it would be inappropriate to intervene. His rights as a ratified participant have been revoked, merely due to the sensitivity of the content of the exchange.

As we’ll see in detail in section 3.3, Clark and Schaefer (1992) interpret ratification within the grounding framework, in terms of the mutual responsibilities the participants have towards one another for ensuring mutual understanding. This, in turn, comes down to the right to speak, e.g. request repair, ask questions, reject etc. In summary, the right to tailor the course of the conversation to one’s own understanding needs. Now, if we are to understand the dynamics of shared-context (when is context shared?) in multi-party dialogue, which is what grounding theory needs to describe, this is a justified interpretation of ratification. We will later discuss some conditions under which the right to speak (strongest form of ratification) can fail to guarantee mutual understanding (e.g. effect of group size). It is one thing having a right but another to actually be able to practice that right to an adequate degree.

3.2.9 Participation as a demonstrative social role

In this section, we will briefly consider how each participant role, incurs a set of normative responsibilities or a certain pattern of (normatively) expected behaviour by its incumbent: “Participation is a demonstrative social role, where each kind of participant role requires a particular kind of appropriate display/behaviour by its incumbent” (Levinson, 1988). A theoretically adequate characterisation of participant roles cannot do without a predictive description of the participants’ behaviours associated with each role.

In this regard, Goodwin (1984) shows how participants organise themselves in relation to each other through the telling of a story, with particular attention to how telling-specific roles, such as teller (speaker), addressed recipient and non-addressed recipient are displayed and differentiated from each other. For example, suppose *A* and *B* are jointly telling a story to *C*. The story starts shifting into an episode in which *D* had some direct and salient involvement, i.e. he is about to become an indirect recipient (like Karen’s case above). *D* now, organises his behaviour such that his outward display is appropriate when all gaze is directed towards him. I.e. he displays ‘indirect recipientship’.

As sketched previously, Kendon (1970), following Condon’s methodology, performs a micro analysis of the phenomenon of non-conscious behavioural synchrony in a focused multi-party encounter. He found, firstly, that different participants synchronise to different degrees with the speaker, and secondly that these different degrees serve particular purposes, e.g. that they clearly distinguish axial from non-axial participants. In particular, it is found that the addressee of the speaker is in higher synchrony with him, when compared to the unaddressed recipients.

When we consider in later sections, Clark’s attempt to distinguish participant roles, we will see that there too, the roles are distinguished in terms of normative principles, in effect specifying what the appropriate behaviour of an incumbent of a certain role should be.

3.2.10 Assignment of roles is not unilateral

Another important point concerning participant role is that it, too - like any other aspect of dialogue - is jointly negotiated rather than unilaterally assigned.

When *A* is addressing *B*, *B* might not choose to attend in that capacity. This is initially the case in the example given previously from Goodwin (1979), where the speaker’s utterance “I gave up smoking one week ago today, actually”, is divided into three distinct utterance-events.

Initially the speaker is unable to secure an addressee and consequently falls back on another. Only upon the third attempt is the speaker able to secure a direct addressee.

Conversely an unaddressed party may try to attract the attention of the speaker by displays of recipientship (postural changes, head movement, overt backchannels, increased synchrony with speaker etc).

Ratification too as has already been discussed, involves negotiation.

Participant role then, is both for the analyst and the participants themselves, collaboratively assigned.

3.3 Grounding and Participation Framework

As alluded to briefly above, Clark and Schaefer (1992) interpret Goffman's ideas within the grounding model expanded in the previous chapter. This is done in terms of a normative description of the mutual responsibilities that dialogue participants occupying different roles have towards one another for ensuring mutual understanding. These responsibilities are then taken to be indexed by participant role. But how?

Ratified Participants in the above sense, are those who adhere to the following principle:

Principle of Responsibility: In a conversation, the parties to it are each responsible for keeping track of what is said, and for enabling the other parties to keep track of what is said.

In other words, they are responsible for making sure that what is said is grounded. As sketched in the previous chapter, in dyadic situations, the speaker and addressee fulfil this responsibility via direct, moment by moment collaboration between themselves. To enable the parties to keep track of what is said, first and foremost the speaker needs to tailor his contribution so as to accommodate his addressee's understanding, i.e. positive audience design. In addition, he needs to respect the addressee's right to repair. The addressee on the other hand, needs to pay attention and signal non-understanding when it occurs. Also, by providing positive evidence of understanding, she needs to keep the speaker aware that she understands what is being said. As we saw earlier, the grounding/contribution model describes a protocol for such collaboration between speaker and addressee, in dyadic dialogue.

In multi-party settings, the situation is more complex, as there are also non-addressed par-

ticipants among those who are ratified. We will henceforth follow Clark and Schaefer (1992) in calling these ratified non-addressed participants, *side-participants*.

Being a side-participant is clearly different from being a direct addressee of an utterance, in that “some answer is . . . anticipated from them [side-participants], more so than from other ratified participants” (Goffman, 1981), meaning that side-participants have ‘weaker’ responsibilities as to providing moment by moment evidence of understanding for the speaker. The obvious question is, how is a side-participant to an utterance-event supposed to collaborate with the speaker while by definition the speaker does not expect her to respond in any fashion, since he is addressing someone else?

Clark and Schaefer (1992) answer this question by positing another principle:

Principle of Collaboration: Speakers collaborate directly with addressees and only indirectly with side-participants.

This basically means that the speaker does not expect side participants to provide direct evidence of understanding, while both the speaker and side participant remain responsible for ensuring that mutual understanding does take place. So the speaker designs his utterances for both addressees and side participants. He is responsible for their understanding of what he says. The side participant on the other hand needs to “be satisfied with clearing up misunderstandings in natural breaks in their [speaker and addressee’s] talk”. (Clark & Schaefer, 1992)

Under the same interpretation of Goffman, unratified participants (henceforth *overhearers*) are those towards whom the speaker bears no responsibility for ensuring mutual understanding. In other words, audience design can be intentionally or otherwise negative for overhearers. With this in mind, as nicely stated by Clark and Schaefer (1992), speakers can adopt a range of attitudes towards overhearers, depending on their goals:

1. **Indifference:** For any part of what they mean, speakers can be indifferent about whether or not the overhearers can grasp it.
2. **Disclosure:** For any part of what they mean, speakers can design their utterances such that the overhearer can grasp it fully.

3. **Disguisement:** For any part of what they mean, speakers can design their utterances such that the overhearers cannot grasp it and will recognise that they cannot do so.
4. **Concealment:** For any part of what they mean, speakers can design their utterances such that the overhearers will be deceived into thinking it is something that it is not.

Regardless of the attitude taken by a speaker on any occasion towards overhearers, they are expected in the general case to reach a much lower level of understanding of what is said compared to the ratified participants. This is largely because they don't have the right to request repair, or more broadly, alter the course of the conversation to accommodate their own understanding needs. The following example should illustrate this point:

- | | |
|-------|--|
| (3.4) | <ol style="list-style-type: none"> (1) A: Did you inform the head of department of what happened yesterday? (2) B: Yes I did. He said to deal with it ourselves. |
|-------|--|

Suppose O is overhearing A's utterance. In order to understand it as B does, he needs to know what department and what yesterday event A is referring to. Clearly, in the general case, he is not guaranteed to know any of this. We will also see in the next section that even if, at the beginning of a conversation, an overhearer shares all the relevant common ground between the ratified parties, he is still at a disadvantage in terms of understanding as a result of not having the grounding cycle at his disposal.

Naturally, according to the above, we expect direct addressees and side participants to reach a higher level of mutual understanding with the speaker, in comparison to overhearers. What follows is a description of a series of experiments carried out by Clark and co-workers, which provide empirical evidence that this is the case.

3.4 Difference in participant role leads to systematically different patterns of common ground

The empirical evidence regarding the level of understanding achieved by participants in different roles, comes from the Tangram experiments, carried out by Clark and co-workers, mentioned

briefly in the previous chapter. The essence of these tasks is that on each trial one person, the Director, describes a series of Tangram figures so that another, the Matcher can identify them. If the same figure recurs on a number of trials the director and matcher quickly converge on some concise definite or nominal description for the figure (recall contraction of referring expressions discussed in the previous chapter. For example, they might go from “Okay the next one is ... resembles someone that looks like they’re trying to climb stairs. There’s two feet, one is way above the other and–” on trial one to “Um, stair climber” on trial six (Wilkes-Gibbs & Clark, 1992).

3.4.1 Addressees and Overhearers

Schober and Clark (1989) carry out the tangram experiment described above while recording the whole conversation between the directors and matchers. Overhearers in this experiment were those who played no direct part in the conversation between the director and matcher but who listened to these conversations on tape, some with the option to pause it and some not. The results were compared in terms of the accuracy of Tangram figure arrangements produced by the matcher (direct addressees) compared to those produced by the overhearers. The matchers were significantly more accurate than the overhearers in arranging the figures. Moreover, those overhearers who had the option to pause the recordings did not perform any better than those who didn’t.

3.4.2 Side participants and Overhearers:

Wilkes-Gibbs and Clark (1992) carried out a two phase variation on the Tangram task. The first phase has two conditions. In one condition, an additional ‘silent’ side-participant (*SP*) sits next to the director. In the other an ‘omniscient overhearer’ (*OO*) follows all of phase 1 on video but is not co-present in the room. In both conditions, the director is aware of the additional participants and they are both able to see each figure as the director describes it. In the second phase the *SP* or *OO* take on the role of matcher for another six trials. The director and former *SP* pairs are quicker, use fewer words and produce more definite/nominal description types than the director and former *OO* pairs; despite the fact that the *SP* and *OO* ostensibly have the same prior information.

It is important to see that any such empirical evidence is really more than a confirmation

of the intuition that overhearers understand less because they lack the information shared between the parties involved at the beginning of a conversation (like the example given above). At the beginning of any of these experiments all those involved (including overhearers) have equal relevant common ground between themselves. Instead these experiments show that direct interaction/collaboration plays a vital role in language understanding and coordination of content in dialogue. Interestingly, these are the experiments that provide evidence directly against the autonomous (vs. Collaborative) view of language understanding. Thus, evidence for the grounding model, elaborated in the previous chapter, comes from contrasting levels of understanding achieved by participants in different roles, in multi-party dialogues.

3.4.3 Side participants and Direct Addressees: Lack of direct empirical evidence

What empirical evidence is there, that side-participants ground to a level comparable with direct addressees despite the fact that by definition they do not collaborate as actively with the speaker as the direct addressee does? Is the mutual understanding between speaker and side-participant as high as that between speaker and addressee? In a review article, Branigan (2006) mentions that there isn't yet any empirical evidence that addressees and side-participants differ in the accumulation of common ground. She also points out that "whereas speakers take silence from their addressees (or, the absence of any overt feedback) as evidence of understanding, it is not clear that the same is true of silence from side-participants".¹

Note that we also need to distinguish between the levels of grounding that side-participants can reach. In the absence of overt feedback, are they taken to have *accepted* what was said between speaker and addressee (tacit consent/acquiescence)? Or are they merely assumed to have understood it? This is consequential for most models, where the effects of a dialogue are expressed in terms of commitments incurred on the part of each participant. Is the side-participant committed to something they did not provide any feedback on?

As described in the previous section, there is clear evidence that side participants reach a higher level of understanding than overhearers. However, the evidence is inconclusive about the side-participant/direct-addressee contrast. This latter distinction relates to participant status with respect to the same utterances (recall the utterance-event/speech-event distinction). The experimental device of two task phases in Wilkes-Gibbs and Clark (1992), effectively breaks the

¹Intuitively, silence from side participants is stronger evidence than from direct addressees since side participants, by definition, can't really provide any evidence while another dyad is engaged in talk.

interaction into two successive conversations where a direct comparison of side-participants and direct-addressees with respect to the same exchanges is not made. The closest approximation is the comparison of the last trial of phase one and the first trial of phase 2 but this is equivocal. The director/side-participant pairs are slower and use more words than the original director-matcher pairs but do make the same number of definite/nominal references. The task situation is also unusual in that in phase one the side-participant is positioned beside the director and opposite the matcher. The participants mutually know that the side-participant has direct visual access to the actual referents of the referring expressions whereas the matcher does not. Arguably this gives the side-participant an unusually high degree of access to the common ground between the director and matcher, whereas in natural conversation this is rarely the case for side-participants.

Below we will explore how this lack of empirical evidence regarding side-participant understanding as compared with direct-addressees, might be highly consequential for the characterisation of grounding in small groups, and why it is therefore crucial that an analysis to this effect should be carried out. This is taken up in chapter 4.

3.4.4 Small group discussions as dyadic dialogue with overhearers?

In their analysis of grounding processes in small groups and how this differs from large groups (which will be expanded in some detail later on), Carletta, Anderson, and Garrod (2002) conclude that small group discussions are like dyadic dialogue, employing the same mechanisms for grounding, but with process loss on the part of peripheral participants (those who are not at that moment, directly engaged in talk) due to material which they merely overheard. But we should see that, particularly in smaller groups, we are really dealing with dyadic dialogues with side participants rather than overhearers (Carletta et al. (2002) do note this but seem to overlook it too easily). As elaborated above, there is as yet no direct empirical evidence that side participants are at a disadvantage in terms of grounding what is said between speaker and addressee. Indeed if it is the case that in certain small groups, side participants ground to a level compared to the dyad directly engaged in talk, then a (general) characterisation of multi-party dialogue as dyadic dialogue with overhearers would be invalidated. It might well be the case, that the side participant understands fully what goes on between speaker and addressee and that they in turn, might assume in the absence/withholding of repair on the part of side-participants, that they have understood them. If this is true then there will be *no process loss*, as the side participant would

have grounded the material to the same level as the dyad engaged in talk. There would really be no ‘overhearing’ involved. The question of course, is to identify where or under what conditions, this may or may not be the case. This is precisely what is at issue in chapters 4 and 5.

3.4.5 Do small groups attempt to establish common ground among all the members?

Moreover, Carletta et al. (2002) describe a map-task experiment in which there are three participants. They find that the group established common ground among all the members even when there was no need to. This leads them to generalise this result to all small group interactions. However, it does not seem overly strange that they got the results they did. They had two route givers with one route follower. The common ground that they “did not need to” establish, is claimed to have been largely between the two route givers. But the two route givers had a number of discrepancies between their maps, and hence would use different ‘situation models’ to describe the routes. In order for them to be able to collaboratively guide the follower, they needed to align themselves (consider the confusion that the follower would have to bear otherwise. The ‘optimal’ case here is where one route giver says nothing and the conversation becomes dyadic). It is the disparity between the route-followers’ maps which is generating that extra talk. But what if the two route givers had the same map? Wouldn’t we, in that case be inclined to say that route giver A would be able to ‘follow up’ on route giver B’s descriptions without having to negotiate anything with A directly? Would we still get the ‘unnecessary’ extra talk? Thus, it seems wrong to claim that the participants “didn’t need” to establish common ground between all the members, given that the experimental conditions of the task were effectively forcing this. We should, at least, acknowledge that this is not generalisable to all small groups. As noted above, we need to characterise the situations in which common ground is explicitly negotiated and those in which it is tacitly assumed without any overt feedback (chapter 5).

3.5 Collaborative responsibilities

Recall the overall pragmatic problem (section 3.2.7). For an utterance-event u , with a set of direct addressees, and a set of side-participants, we need to characterise the set of responsibilities that hold for each participant in either the role of an addressee or a side-participant, locally with respect to u .

The Principle of Responsibility (section 3.3) states that every ratified participant is responsible for making sure that what is said is understood by all other ratified participants. But this is

just a normative description of what should happen, not a systematic description of *how* it happens. In other words, Clark and Schaefer (1992) do not specify exactly what mutual grounding responsibilities hold at the local, utterance-event level between participants in different roles. In this section we will attempt to raise some related questions.

As stated in section 3.3, the Principle of Collaboration is what discriminates side-participant/speaker from addressee/speaker responsibilities. However, this is really as far as the authors go in trying to do this. It remains a question therein:

1. what exactly constitutes a speaker's indirect collaborative responsibilities towards side-participants?
2. and, likewise, what, if anything, constitutes a side-participant's collaborative responsibilities towards the speaker and addressee?

With regards to the first question, the answer could lie in positive Audience Design. But we already know that even in dyadic dialogue this can fail (recall Keysar and colleagues' criticisms of Grice's cooperativity principle, summarised in section 2.3.7). Indeed, Clark and Brennan (1991) state that "Speakers often realise they just don't know enough about their interlocutor to design a proper utterance. So they are forced to issue an improper utterance instead". In multi-party situations, it would be even more difficult for a speaker to design her utterances for all her interlocutors (side participants as well as direct addressees). In fact it seems that positive audience design for side participants is often quite weak.

At the beginning of any ordinary multi-party dialogue, at least the personal common ground between each pair of participants is unique. Taking it that utterances are interpreted against common ground, what counts as a felicitous utterance by A addressed at B would clearly not in the general case be felicitous if addressed at C. Indeed, C as a side-participant might not even immediately understand what A is telling B. But being a side-participant means that he can request repair. We are faced with an immediate problem here: To what degree do speakers consider side-participants in the design of their utterances? According to Clark and Schaefer (1992), when A addresses B with C as a side-participant, he intends both B and C to understand. But it seems that this intention is only manifested prospectively in a speaker's willingness to repair his utterances upon the request of a side-participant:

- (3.5)
- (1) A: I was trying to debug a programme I spent 3 hours writing. It took me about two more days just to finish the debugging. [addressed at B and C]
 - (2) B: Yeah . . . like John's project. Do you remember? [addressed at A]
 - (3) A: oh yeah . . . Quite vividly actually.
 - (4) C: John? [addressed at B]
 - (5) B: Oh sorry erm John is a friend of ours.
 - (6) C: oh right. Did the same thing happen to him?

Clearly, B's first utterance would be infelicitous if it was addressed at C. The point is that in term of audience design, a speaker really only has strong commitments towards the addressee. He always tries to design his utterances at least against his common ground with the addressee. In the above example, B would have to turn to C during the process of answering A's question, to say what 'my project' refers to. In fact it would be wrong to claim that B is even considering C at the point of answering A's question. Of course, since C is ratified, B is more than happy to provide further information upon request.

However, we should also note that weaker commitments towards side-participants aren't necessarily to do with lack of knowledge by the side-participant as in example 3.5. It may be the opposite. Grice's Maxim of Quantity for example, really still holds only for the addressee. For instance, while it is no problem to address *u* at *A*, while a side-participant *B* already knows about the content of *u*, this would not work the other way around. Recall the example from Goodwin (1979), given in section 3.2.4, on how a single turn can contain more than one utterance-event. The speaker, in his attempt to secure an addressee, has to redesign his utterance on the fly as his gaze moves from one participant to the other, as the same utterance, as it stands, would not be appropriate for who was until now a side-participant but is now an addressee. Hence he redesigns the utterance to include something that this present addressee does not know (adding "one week ago today actually" to "I gave up smoking").

It seems that, strictly speaking, speakers collaborate with side-participants largely by respecting their right to repair/speak.

Regarding the second question above, about side-participant responsibilities with respect to something uttered by the speaker, what level of evidence of understanding, if any, do they need to provide? Continued attention? Withholding of repair? If the latter, for how long? Do side participants and direct addressees have any responsibilities towards each other with respect to something uttered by the speaker? We address some of these questions in chapter 5.

3.6 The pragmatic force of an utterance-event and the problem of multiple interpretations

In section 3.2.7, we sketched the overall problem that any theory of multi-party dialogue context would have to deal with. Namely, given an utterance-event u and participant roles with respect to it, how does u change the context for each of the participants. This is a question which Speech Act Theory attempts to answer. But apart from the various problems the theory already faces in dyadic situations, we run into fundamental difficulties in generalising the theory to cover multi-party phenomena. These difficulties are due mainly to the fact that the pragmatic force of an utterance-event is relative to the participant role of the participants involved in the event. Let's see how this is so.

For reception roles, if A makes a request of B with C present as a non-addressed party, then clearly this places an obligation on B to do what A asks or issue a rejection. But no such obligation is placed on C. Similarly a question only commits the addressee to answer.

As for production roles, a promise commits the speaker. But if the speaker is merely an *animator* of the message, the promise should only commit the *principal*.

Clark and Carlson (1982) attempt to solve this problem for the reception role set. Very briefly they do this by adding another class of speech acts to the originals (request, inform, promise etc). They call these *participant-directed* acts and the original acts, *addressee-directed* acts. The proposal is that in any addressee directed act, the speaker is in addition to the act being performed upon the addressee, performing corresponding *informative* acts towards the non-addressed participants. This is intuitive since A above is doing more than request something from B. He is also informing C that he is making this request of B.

However, we should see that issues of participant status alone, do not exhaust the problem of interpretation relative to shared-context/common ground. We sketched this problem in section 2.3.7 of the previous chapter. The problem is that in multi-party dialogues, for any utterance uttered by A, the role of common ground (in each case between A and some hearer), in the interpretation of the utterance by a hearer, implies that relevant differences (with respect to current conversational topic) in what A already shares with each hearer, will result in different interpretations by the corresponding hearers, of the same utterance. This is particularly true of elliptical utterances.

For example, Sacks, Jefferson, and Schegloff (1995) state that if A addresses B with C as an

unaddressed recipient, the utterance can do something to B that is different from what it does to C. For example if A flirts with B, A might be teasing C. Although this is possible only in this specific arrangement of the participation framework, such differences in interpretation are really to do with the prior shared-context between A and C, and what A assumes he knows about the shared-context between B and C, and cannot be otherwise explained.

A more radical case is that of ‘multiple recipient design’, that is, where “the speaker attempts to tell, within one and the same turn, a proposition A to the therapist, but non-A to his co-present wife” (Muller, 1997). This is part of Muller’s comments on Lonardi and Viaro’s (1990) work on interviews between therapists and their patients. Here, again, this is clearly not in and of itself a matter of the distinction between addressed and non-addressed participants

Another example is as follows:

A and B go food shopping. They have decided to buy milk and some other stuff. They come back home. A realises that they have forgotten to buy milk. In the presence of C (who is a fully ratified participant), he points to an empty milk bottle on the table:

- | | |
|-------|---|
| (3.6) | (1) A: oh shit, milk!!
(2) B: oops. [resolves as “oh yeah we forgot to buy milk”]
(3) C: what? what’s wrong with milk?
(4) A: no I meant we forgot to buy milk.
(5) C: oh. haha |
|-------|---|

For such a situation to be possible at all, there has to be some form of pragmatic displacement - difference in interpretation - occurring between the two readings of an elliptical move, owing to the difference between the contexts against which they are interpreted. This demonstrates how interpretation of utterances in dialogue, depends upon the common ground between every pair of participants - in each case a speaker and a recipient.

In chapter 6 we will return to this problem, where we present conditions under which shared contexts can systematically diverge in one and the same conversation, resulting, as above, in a systematic disparity between the interpretations of the same utterance, by different participants.

3.7 Turn-taking in multi-party dialogues

Models of turn-taking attempt to capture and make explicit the set of norms by which conversationalists manage to take turns to talk in an orderly manner. Indeed, turn-taking is not limited to

the activity of talking, but is broadly applicable to activities in which there is a single resource to be shared by several people in some orderly fashion. The resource to be shared in the socially organised activity of talking is the *floor*. Thus, accounts of turn-taking in verbal interaction need to determine, given the history of the conversation, who will hold the floor next, i.e. who speaks next. Naturally, this is to be determined according to the state of the floor at the moment, i.e. the current speaker, and what he is saying.

3.7.1 The Simplest Systematics

The most prominent account of turn-taking in verbal interaction is that developed by Sacks et al. (1974), within the CA tradition. This is widely referred to as the ‘simplest systematics’ of turn-taking. We’ll see how such a label is justified below. The most general rule/norm enforced by this system is ‘one speaker at a time’. Of course this is not a matter of choice on the part of the analyst, but is empirically supported by the observation that overlaps in talk are quite short in length (though not infrequent). This shows an overwhelming preference by the people engaged in talk, for the rule above.

On this account, speaker transition can occur at moments referred to as *transition-relevance places* (TRP). These occur at the end of turn-constructive-units (TCU), which are instances of various unit types of a language, e.g. phrase, clause, sentence. At the end of any such unit - at the TRPs which are discrete moments in the course of a turn at talk - the following rules apply (here quite roughly stated):

1. If the turn-so-far involves a next speaker selection, on the part of the current speaker, then the selected next speaker has exclusive right to the floor, i.e. he is obliged to speak next.
2. If 1 does not hold, there is free, equal competition for the floor, meaning that all the rest of the speakers can self-select. The one who speaks first wins the floor.
3. If nobody self-selects - 2 does not hold - the speaker may, but is not obliged to continue.

As evident in the rules above, this model is developed to be independent of all contextual parameters including topical, social, circumstantial, and most importantly here, number of participants and thus applicable to any verbal conversational setting including multi-party interactions.

The simplest systematics has not remained uncriticised over the years. Below, we will consider some of these criticisms.

3.7.2 Some Critiques

There are three main aspects of the turn taking system outlined above, which have been criticised.

These are:

1. The rule on which the simplest systematics is based, namely ‘One speaker at a time’, need not be a necessary condition for communication in general.
2. The simplest systematics is not and need not be cross-culturally valid.
3. The CA system for turn-taking not unlike other CA formulations, is supposedly intended as a set of prescriptive rules for people to adhere to rather than a descriptive account of the succession of turns in verbal interaction.

As regards the first, Edelsky (1981) points out that instances of simultaneous talk need neither always be brief nor infelicitous. Spelke, Hirst, and Neisser (1976) had shown earlier that people can in fact process multiple simultaneous messages coming from different sources, or at least understand the gist of those messages, provided that they weren’t overly long or complex.

We should therefore expect that, in certain communities, speech overlaps are more common and less brief than others: “there is at least one speech community where naturally occurring simultaneous talk is frequent, expected and processed.” (Reisman, 1977). On this view, such systematics for turn-taking are not valid cross-culturally or even across social groups within the same community. For example, “Italians interrupt one another. Everybody gets all excited and tries to make his views prevail by preventing the other from speaking Americans speak in turns. (It is no accident that the pragmatic theory of ‘conversation turns’ originated in the United States. Italian researchers who write articles about this matter treat it as an excavation from Mars.)” (Eco, 1986) quoted by O’Connell, Kowal, and Kaltenbacher (1990).

The third point above, as sketched in the beginning of the previous chapter, is in general a source of contention with respect to the overall practice of the CA community. Power and Martello (1986) point out that the authors of the simplest systematics:

assume too easily that a regularity in social behaviour must be due to a social convention specifically prescribing that regularity. Observing that the first person to speak up is usually allowed to continue, or that addressed questions are usually immediately answered by the addressee, they assume without discussion that these

regularities are due to specific turn-taking conventions In making this criticism we are assuming, of course, that the rules given by SSJ [simplest systematics] are supposed to represent a cultural convention and not merely a statistical regularity.

This controversy is due mainly to the philosophical contrast between prescriptive and descriptive rules. Prescriptive rules are essentially those, which people (should) follow due to social conventions. Descriptive ones on the other hand are only to predictively capture some set of social data without much concern for the cause - why the data is as it is. But what difference does this make so long as these 'rules' do not fail to predict/capture social behaviour in a certain predetermined community? It is perhaps relevant here to point out that in my speech community (Farsi speaking, Iranian) rules of turn-taking as delineated by the simplest systematics certainly hold as prescriptive. A child is almost always instructed by her/his parents not to interrupt, and particularly in case an elderly or a person with social status holds the floor, not to speak unless spoken to. It is otherwise considered 'rude'.

In conclusion to this section, we should note that none of the criticisms above invalidate the systematics so long as we restrict ourselves to the English speaking community. The main point is that turn-taking in conversation follows fairly systematic principles, albeit potentially different ones in different speech communities.

3.7.3 Small group discussions are comprised of a series of pairwise discussions

As we just discussed, the simplest systematics operates independently of the number of participants present. However, it is in essence dyadic. It operates between two speakers, namely current and next. It arranges but two speakers at a time. Consequently the model projects a dyadic 'in-response-to' relation between utterances.

In the floor model developed in Parker (1988), the discussion is in a 'floor-state' if the current utterance falls within an A-B-A pattern of turn-taking, where A and B are participants. This was applied to data from focused group discussions between 4 persons. The significant finding is that the discussion is, by a large margin, most likely to be in a floor-state. Thus, dyadic turn sequences are shown to be preferred. Even if such patterns are broken, new ones are quickly established. The same floor model has also been applied to groups of 6 (Stasser & Taylor, 1991). The likelihood of the discussions being in a floor state in these groups is shown to be still the most likely, though less so than in the 4-person groups in Parker's original study.

Most models of turn-taking, then, seem to suggest that small group discussions are primarily comprised of a series of pair-wise discussions, that are interrupted by another participant to establish new pairings. But note that it is conceivable that such conversations aren't really dyadic. This would happen, as we suggested in section 3.4.4, if a side-participant reaches the same level understanding as speaker and addressee, in which case the claim that multi-party dialogue is comprised of a series of pair-wise discussions might be misleading, if not wrong.

Let's now turn to how turn-taking patterns and hence grounding might be affected by group size, and status differences among the participants.

3.7.4 Effect of Group Size on Turn-taking/Grounding

We see above that as group size increases the likelihood of dyadic patterns of turn-taking drops. Floor-states do not arise as often in large groups. Dyads have less momentum. In a more recent study, Fay, Garrod, and Carletta (2000) show that in small 5-person groups, the discussion exhibits dialogic properties, where the participants' opinions about what went on in the discussion are most influenced by the people with whom they interacted most. Indeed if small groups behaved like dialogue, this is what grounding theory would predict (mutual understanding is established among those who have interacted with each other and not others). In contrast, in large 10-person groups, opinions are most primarily affected by who spoke the most - the dominant speaker, i.e. large groups exhibit more monologic characteristics. Larger groups are thus, more likely to behave like monologue where communication takes place between a sender and a receiver who process language signals autonomously, hence the claim that grounding is not at work in these groups and that instead they behave rather like a series of lectures where the next lecturer is chosen from the audience. Although as yet unstudied in this respect, one would also expect a much lower degree of pragmatic cohesiveness (less cross-turn context dependency) in such discussions. We'll say more about this issue below.

In cocktail party settings where higher degrees of interactivity and involvement are preferred, larger groups are systematically avoided. Hare (1952) notices that large unfocused groups of 12 in which people can leave and join the discussion freely tend to split into cliques more often than groups of 5 with the same property. Also, Homans (1950) observed that groups in such settings tend to split roughly when they reach a size of five. So it seems generally impossible to maintain a large group discussion without other social, organisational constraints - where the group 'have to' stay together for a period of time.

But why is all this happening, one might ask. The reason is mainly the exponentially increasing number of possible pairings as groups get large. Also how people manage turn-taking in face-to-face dialogue, relies heavily on visual monitoring between the parties. For example, the participants to whom the gaze of the speaker is directed at the end of a turn are more likely to speak next (Kalma, 1992). Moreover, there are many speaker behaviours which enable hearers to predict when the speaker is coming to the end of her utterance (Duncan, 1972). All of these mechanisms and more, become more difficult if not impossible as group size increases. A person's attempt to take a turn at talk can easily go unnoticed and this affects group behaviour. Boden (1994) observes that participants in larger groups produce fewer backchannel utterances than those in smaller ones.

Following what was said above about the study in Fay et al. (2000), Carletta et al. (2002) propose that there is in fact a model split in terms of behaviour between groups of size seven and eight. It is claimed that groups of size seven or fewer "appear to employ a straightforward extension of the grounding processes that are well understood from previous work on dialogue". These discussions are also said to behave like dyadic dialogue with overhearers (though recall our discussion in section 3.4.3, meaning that the mechanisms employed for grounding remain the same as dyadic dialogue, where there is process loss because of material which is merely overheard). Groups of 8 or more on the other hand (like the groups 10 in Fay et al. (2000)), should behave, according to the paper, like monologue.

This result regarding the split between groups of 7 and 8, comes from observing that the distribution of the length of turns in groups of 7 or less in the data analysed by Carletta et al. (2002), are statistically indistinguishable. The same is true of groups of 8 or more, while the difference between sizes seven and eight was sudden rather than gradual. Turns were suddenly longer in size 8 groups, where interactive discussions (where utterances can clearly be thought of as 'responses' to other utterances) among pairs were shorter and also a lot less frequent. Hence the claim: "small and large groups appear to operate according to radically different processes" (Carletta et al., 2002).

However, the contrast between monologue and dialogue seems unclear. Certainly the degree of interactivity is the determining factor here. But what does it mean exactly? Is the observation that utterances are largely broadcast to the whole group and that they are longer, necessarily an indicator of smaller degrees of interactivity? Perhaps a good measure of interactivity, as men-

tioned briefly above, is the degree of pragmatic cohesiveness between and across turns. Ellipsis, co-reference and in general the degree to which the meaning of a speaker's utterance depends on others' utterances - dependence upon some notion of shared-context - could be empirically investigated in order to shed more light on such results. It would be expected that there should also be a model split in terms of this latter measure between groups of size seven and eight. Carletta et al. (2002) however, do not provide any evidence of this form.

3.7.5 Effects of uneven power distribution on turn-taking/grounding

Status² has long been known to affect communication. Group communications, except perhaps in small groups, are centralised around group leaders (Hare, 1981). This in turn directly affects patterns of turn-taking and consequently grounding. Carletta, Garrod, and Fraser-Krauss (1998) reveal that regardless of whether there is a powerful individual present, small group discussions are still primarily comprised of a series of pairwise discussions. The difference lies in the pairings established in each case. When decision making authority is placed upon one individual the same pairs tend to recur, whereas if authority is placed on the group as a whole, the pairings tend to be more evenly distributed.

Apart from the group size factor, the study in Carletta et al. (2002) also compares traditional with autonomous work groups. The former type, contain a powerful individual leading the conversation. In the latter, authority is uniformly distributed among the participants. In traditional work groups, pairings recur due to the presence of the powerful individual. These were shown to behave like large groups as described in the previous section. Moreover, group size and individual status are known to interact. It has long been known that in equal-status, large groups one person will come to dominate the discussion (Bales, Strodtbeck, Mills, & Roseborough, 1951). In other words, even if there isn't a high-status individual, one will emerge if the group is large enough.

Thus, even though grounding tools seem to remain the same, the resulting states of understanding are skewed and heterogeneous, privileging powerful individuals in terms of understanding and being understood. Autonomous and small work groups are for this reason considered best if true group collaboration/decision making is to be achieved.

²We are not referring to participant status here, rather social status, e.g. a powerful individual or a group leader

3.8 Alignment and priming in multi-party dialogue

We pointed out, in section 3.4, how differences in participant role lead to systematically different patterns of common ground. But what are the mechanisms that account for these differences? The grounding model would explain these differences, in terms of the opportunities and obligations, or the lack thereof, that participants in different roles have to engage in, in the grounding cycle. On this account, peripheral participants reach a lower level of understanding because they don't have grounding as a resource.

What would the Interactive Alignment Model (Pickering & Garrod, 2004) have to say about these differences? As sketched previously, the interactive alignment model states that successful communication is achieved via automatic and resource-free priming mechanisms at different linguistic levels. But priming mechanisms, or the coupling of production/comprehension or input/output, should operate independently of participant role, and hence the model would have difficulty explaining these differences.

However, Carletta et al. (1998) suggest that the same priming mechanisms which operate in dyadic dialogue are also at work in multi-party dialogue. Their key point is that peripheral participants don't have the opportunity to address problems in interpreting the primary participants' utterances, and hence will be less coordinated with the primary participants because they are, in effect, exposed to fewer instances of each input.

Moreover, priming seems to be sensitive to participant role also at a local level. For example, Branigan, Pickering, and McLean (2005) found that side participants as well as direct addressees align syntactically with a previous speaker, but to a lesser degree. They found that prior side participants are less likely to align with the previous speaker than prior addressees. Furthermore, various socio-linguistic studies have shown that speakers are less sensitive to side-participants than they are to their addressees, with respect to e.g. lexicon and phonology. In other words, perceived attributes in an addressee's immediate behaviour have a greater impact upon the speaker's subsequent behaviour.

Nevertheless, Healey and Mills (2006) provide evidence of both local and global semantic coordination phenomena that are not captured by priming mechanisms. The paper reports an experiment that investigates the effects of: a) apparent origin of an utterance (primary vs. peripheral participants) and b) prior exposure (priming) on the development of inter-speaker semantic coordination (alignment at the level of the situation model). In these experiments, random

computer-generated probe questions get systematically different answers depending on whether the apparent origin of the probe was a peripheral (passive, silent but ratified bystanders) or a primary participant. Therefore differences in participant role - or levels of participation - lead to different levels of semantic coordination. They argue that these differences are due to the opportunities that direct interaction creates for addressing differences in interpretation.

Still to be investigated is whether their results would persist if the peripheral participant was actively involved in the interaction between the other two, and hence was only a side participant locally, on and off, during certain phases of the dialogue. If it is indeed the imbalance in degrees of participation/involvement which is causing the differences in levels of coordination, then we would expect their results not to persist. If so, then we would need to recognise degrees of peripherality, i.e. we need to distinguish between silent, passive bystanders and active, responsible side-participants, independently that is, of ratification. But what exactly is the difference here? This is an issue that we take up in detail in chapter 5.

3.9 Repair in multi-party dialogue

Returning to our discussion of repair from the previous chapter, let's begin with some questions in this regard which arise in the context of multi-party dialogues:

1. How do side-participants who have had to stay silent during a stretch of dyadic interaction between another two participants, signal problems in understanding what they said to each other, i.e. initiate repair? Is the structure of such repair initiations different from those of the primary participants themselves?
2. Do side participants perform other-repair upon what the active parties say to one another? What is the structure of such third person repairs? Under what circumstances can and do they occur smoothly?
3. Can more than one person initiate repair on the same trouble source turn? Is the response tailored to the specific participant?
4. Who responds in a multi-party setting to the repair initiation? Is it always the speaker of the trouble source?

Regarding question 1 above, as mentioned, Clark and Schaefer (1992) states that a side-participant often has to wait for speaker and addressee to carry out their presentation and ac-

ceptance phases, before they can interject. Therefore the side participant repair initiation may be ‘displaced’ from the trouble source, as a result of the third party having to wait for a gap in the active parties’ talk. If this is true, such repair initiations can’t rely on adjacency of the trouble source. The question arises therefore, as to whether there is any extra effort that the side-participant would need to invest, in order to build their repair/repair initiation as ‘next to’ the trouble source turn? Would they need to be less elliptical? This problem ties in more closely with the more general issue of sequential coherence in multi-party dialogues. We will therefore defer further discussion until section 3.10.

Concerning question 2, to the author’s knowledge there hasn’t been a systematic study of such side participant repairs. Suffice it to say that answering it requires an extensive corpus analysis of side-participant repairs in multi-party dialogues. Intuitively side participant other-repairs could occur, as a misunderstanding might be revealed to the side participant, which might have gone unnoticed by the active parties themselves - as a result of higher degrees of alignment of perspectives between the side participant and one of the primary participants as compared to that between the primary participants themselves. But they would probably be very rare, as other-repairs in dyadic dialogue are generally dispreferred. Side-participant other-repairs are likely to be even more dispreferred, as the side participant isn’t even the addressee of the trouble source. However, they are conceivable if participants have a very close and less formal relationship, in which case such repairs would seem less face-threatening.

Egbert (1997), in line with the CA tradition, performs a qualitative analysis of other-initiated repairs in multi-party conversations. She shows that, with regards to our 3rd question above, more than one person can initiate repair on a single trouble source turn. However, in all the transcripts used, the repair initiation only receives one answer, i.e. the speaker does not repair her utterance more than once for each repair initiator. Egbert argues that such redundancy has a specific interactional function, namely that of a momentary display of affiliation by the second repair initiator to the first. In other words this is one way in which parties/collectivities (a notion discussed at length in chapter 5) can be projected in multi-party conversation, in this case the collection of people who had a problem understanding the trouble-source.

Nevertheless, we should see that it is conceivable for repair initiations on the same trouble source, to get different answers from its speaker, owing to the differences between the contexts shared - common ground - between the speaker and each of the repair initiators.

Egbert (1997) also shows that it is more often than not the speaker of the trouble source who responds to repair. However, there are cases where another speaker responds before the speaker of the trouble source. In such cases the original speaker's reply is marked, in a way that shows that such practice is a deviation from the norm, and hence 'inappropriate'. This shows that, in multi-party dialogues, the responsibility to address any trouble in others' understanding of an utterance, remains with the speaker.

3.10 Structuring of multi-party dialogue

3.10.1 Exchanges

Dyadic dialogue is usually structured into the following hierarchy of units:

- *Move*: This is the smallest unit, corresponding usually to a monologic speech-act.
- *Turn*: Turns which are still monologic, are comprised of a number of moves.
- *Exchange*: Exchanges are defined as the smallest pragmatic units to be produced by at least two different speakers, comprised of exactly one initiating and one reacting move. E.g. see Traum (1994) for an analysis of dyadic exchanges in terms of forward-looking and backward-looking Conversational Acts.

In multi-party dialogue however, the question of the completeness of an exchange is posed in different terms. Not unlike dyadic dialogue, the initiating move can be comprised of a number of utterance-events, produced by different participants. However, an exchange in multi-party dialogue is maximally comprised of one initiating move to which all the addressed and non-addressed participants (when speaker selection rules are violated) may then react (produce second pair-parts). Delamotte-Legrand (1995) for example, studies the division of exchanges in multi-party dialogues between 9 pre-adolescent children and found that exchanges with 3-5 and 4-8 moves were dominant.

To illustrate the point, a question can get a number of different answers from different hearers in different roles, or an assertion can be accepted or rejected by all the ratified participants. Moreover, all this can be done elliptically as long as the antecedent of the ellipsis is 'still salient' in context. Consider the following example:

- (3.7) | (1) A: Have you done your home works?
 (2) B: Yes
 (3) C: Yes
 (4) D: No

The above is an example of a 4-move exchange. The reason why *D* can produce a ‘No’ as he does, is because the question is still salient at that point. We are still ‘within’ the same exchange. *D*’s answer is as local to *A*’s question as *B*’s. Now suppose that *B* had said something to move the conversation on such that the exchange would be cut short, *truncated*. It seems, then, that *C* or *D* would not be able to produce such elliptical responses to the question, since what would then be salient is *B*’s extended move as initiating a new exchange while truncating the previous one. But what are the criteria for truncation? E.g. how is it that an elliptical response from *B* above does not truncate the exchange whereas an extended utterance might?

Such cases give rise to problems of distance and context locality, since ellipses - at least elliptical answers and grounding acts - are presumed to have local antecedents. But what should the notion of locality be here considering that it should not distinguish between *D*’s and *B*’s utterances? As we pointed out in section 3.2.5, this cannot be number of turns or time elapsed since the antecedent. An alternative notion of locality/saliency is developed and tested experimentally in chapters 6 and 7.

3.10.2 Multiplicity of Conditionally Relevant Contributions

Consider dialogue 3.7 again. Suppose that *D*’s utterance is this time addressed at *C* instead of *A*. Then we’d be strongly inclined to take *D*’s utterance as a rejection of *C*’s. In other words *D*’s utterance ‘picks’ a different antecedent. The ellipsis is then resolved in the two cases respectively as follows: “No, I haven’t done my homework” and “No, you [*C*] haven’t done your homework”.

Put in Conversational Analytic terms, we would say that at the end of *C*’s utterance, an answer to *A*’s question and a rejection of *C*’s answer to *A*’s question, are BOTH conditionally relevant. The adjacency pairing structure - according to which the antecedents of ellipses are determined - is contingent upon what exchanges are currently salient, local and active, and therefore selectable as a first pair-part to the current contribution.

So, every utterance initiates a new exchange and is therefore a potential antecedent for subsequent ellipsis. Ambiguity arises when we're still within another non-truncated, salient prior exchange, as with A's question in 3.7.

3.10.3 Context re-raising

Moreover, it appears as though questions of context saliency/locality are really only relevant if we are to determine when and where a certain form of ellipsis is licensed or felicitous. In contrast, utterances which aren't elliptical seem to be possible at any point within a conversation and indeed even if the conversation has been terminated in the sense put forward in Goffman's definition given previously.

Although most turns respond to the immediately preceding talk, speakers can construct turns to address much earlier talk . . . as what their ensuing talk is built as "next to". (Schegloff, 1992)

This is done via re-raising relevant parts of the context - avoiding highly elliptical expressions or in the case of anaphora and definite references, giving further descriptions of the discourse entities referred to - that the speaker already shares with his interlocutor, in order to make the desired contribution - the second pair part to a first which is too far back for elliptical access. But what form do these re-raising take? What parts of the context need to be re-raised and what parts can be safely referred to anaphorically? We will see immediately below how this issue is also directly relevant to the intuition that side-participants are distinguished from addressees in terms of initiative and accessibility of the floor.

As mentioned previously, there is a strong intuition that SP's are different from the primary participants as to how they can access the common ground built up between the primary participants. "they [side-participants] have to be satisfied with clearing up misunderstanding in natural breaks in their [speaker and addressee] talk" Clark and Schaefer (1992). A side-participant often needs to wait for them to carry out their presentation and acceptance phases - and possibly resolve all the subquestions therein - before they can interject. This is even more the case, where there is an uneven power distribution among the members of the group, e.g. if the side-participant has a lower social status and less confidence with respect to what the primary parties are talking about. Thus, anecdotally, it seems that side-participants frequently use techniques for re-raising context

in order to access the prior context, which is at this point 'too far back' for elliptical access. So, on this account, side-participant utterances should be overall less elliptical.

Existing empirical evidence for context re-raising, comes from Purver (2004)'s corpus study of CRs (Clarification Requests or repair initiations). Having produced a taxonomy of the forms that CR's can get, he goes on to show that the further a CR is from its antecedent or trouble-source, the less elliptical its form will be; such that at long distances, CR's take the form of full-blown clarification questions where all that is asked about is contained wholly within the CR itself.

A similar issue arises regarding the interpretability of elliptical utterances relative to topic changes. Do 'new' questions/topics - which do not count as sub-questions of ones introduced previously within a dyadic exchange - make the ones introduced and resolved before, inaccessible as possible antecedents? If so, this situation too necessitates context re-raising. We should already see that this is really the same problem as the one outlined above regarding the truncation of exchanges.

3.11 Transformations to Ginzburg's KoS: A short review of the empirical benchmarks

In the previous chapter we sketched Ginzburg's model of dialogue context. In this section we will review the empirical benchmarks on which the proposed transformations to the model to handle multi-party dialogue, are based. We will not, however, expand here on the the transformations themselves.

The transformations carried out on the original model, proposed by Ginzburg and Fernandez (2005), are based on two empirical benchmarks derived from a study of the distance between Non-Sentential Utterances (NSU) and their antecedents. NSU's are fragments which are meaningless without the background of the conversation, such as a Short Answer which could be a one-word reply to a question (for a taxonomy of these, see Fernandez and Ginzburg (2002)). The antecedent is the utterance from which the meaning of the NSU is recovered, e.g. the question to which a Short Answer is a reply. The distance between the two is measured here in terms of turn/sentence numbers.

The study shows that (1): NSUs used for grounding, namely acknowledgements/acceptances and clarification ellipses have a distance of at most 1 from their antecedents, i.e. the antecedents

are immediately prior to the NSUs. And that (2): long distance short answers were shown to be primarily an effect produced in and by multi-party situations. The corresponding benchmarks developed from these results are:

- **Multilogue Long Distance short answers (MLDSA):** querying protocols for multilogue must license short answers an unbounded number of turns from the original query.
- **Multilogue adjacency of grounding/acceptance (MAG):** assertion and grounding protocols for multilogue should license grounding/clarification/acceptance moves only adjacently to their antecedent utterance.

The model is then transformed such that the result satisfies both of these benchmarks. Let's explore them in turn.

Recall our discussion in the previous section, of completeness of exchanges in multi-party dialogues. The first benchmark is along the same lines. It is perfectly intuitive, since a question in multi-party settings, can have more than one direct addressee, which means that the last participants to take a turn will have produced long distance short answers.

The second benchmark is satisfied by protocols in which *communal acceptance* is enforced. Under this assumption, an utterance is both grounded and accepted by all the ratified participants, as soon as one of them signals acceptance. Intuitively, this is quite curious, as multi-party dialogues of the following sort seem natural:

- **Grounding at distance:** It is perfectly natural for a number (> 1) of direct addressees to consecutively signal their understanding. Also, in multi-party dialogues, it may be that (elliptical) clarification requests can be posed smoothly by side participants at distances of more than 1. Dealing with such instances would require a distributive grounding protocol. Although the authors do note that this may be a problem they do not address it.
- **Discussion at distance:** We need to consider the possibility of long distance rejections or discussion. Again, it is not inconceivable that a side participant should signal their disagreement, at distance, after the primary participants have finished talking about the issue under discussion.

What would go wrong in the model in these cases, is that a question would be popped off QUD 'too early', i.e. while a side-participant could felicitously acknowledge/accept/reject/clarify at turn distances of more than 1, she cannot do so under the induced protocol. Again Ginzburg and Fernandez (2005) do note briefly that such cases would involve some sort of backtracking (both at the level of acceptance and identification/understanding) but don't give any ideas as to what this might involve.

The second benchmark therefore bears closer empirical/corpus examination. This will be carried out in chapter 4.

Chapter 4

Collective states of understanding

4.1 Introduction

As discussed in Chapter 3, Goffman (1981) introduced the distinction between *ratified participants* and *overhearers* in a conversation. The former category is further decomposed into *direct addressees (DA)* and *side participants (SP)* of an utterance. The ratified participants are those who hold certain responsibilities towards each other for ensuring mutual-understanding:

Principle of Responsibility: In a conversation, the parties to it are each responsible for keeping track of what is said, and for enabling the other parties to keep track of what is said. (Clark & Schaefer, 1992)

As we saw in detail in Chapter 2, in dyadic interactions, mutual-understanding or ‘grounding’ is achieved through direct collaboration between the speaker and addressee. The speaker expects the addressee to provide evidence that he is understanding the speaker’s utterance “to criteria sufficient for current purposes” (Clark & Schaefer, 1989). In multi-party conversations on the other hand, the situation is more complex.

For example, if A makes an anaphoric reference to some entity, while addressing B with C present as a side-participant, he intends both B AND C to resolve the reference. However, by definition, the speaker does not collaborate as actively with side-participants. They “have to be satisfied with clearing up misunderstandings in natural breaks in their talk”(Clark & Schaefer,

1992). A *SP* will normally wait until speaker and addressee have carried out their *presentation* and *acceptance* phases, before attempting to rectify any possible misalignment with the speaker. On this account grounding between speaker and direct addressee always takes precedence.

By definition, then, *SPs* and *DAs* give different evidence of grounding of a speaker's utterances; *DAs* respond overtly and directly but *SPs* provide weaker evidence of grounding – primarily continued attention and withholding of repair. Consequently, if we understand level of grounding as being directly contingent upon the level of 'evidence of acceptance' provided then we expect differences in the relative accessibility of the common ground for the different pairs of participants; roughly, Speaker & *DA* > Speaker & *SP* > *SP* & *DA*.

In a review article Branigan (2006) points out that there isn't yet any empirical evidence that *DAs* and *SPs* differ in the accumulation of common ground. Furthermore, as we discussed in section 3.4.3, although the Tangram experiments carried out by Clark and colleagues (Wilkes-Gibbs & Clark, 1992; Schober & Clark, 1989) show clearly that *SPs* and *DAs* reach a higher level of grounding than overhearers, they do not provide direct empirical evidence as regards the relative status of *SPs* and *DAs* with respect to the accumulation of common ground *in the same conversation*. This is because the two task phases in these experiments effectively break the interaction into two consecutive, separate conversations where a comparison between the level of understanding of *DAs* and *SPs*, in the same conversation, at the level of the utterance-event, is not made.

4.1.1 Group discussion as dyadic dialogue with overhearers?

groups . . . behave like pairs of speakers with overhearers . . . dialogue participants try to mutually understand each other, establishing common ground **among all members of a group** . . . so that small group discussions can be seen as [dyadic] dialogue with **process loss** due to the difficulties of fully understanding material which is merely overheard. (Carletta et al., 2002)

Nevertheless, as discussed in section 3.4.4, Carletta et al. (2002)'s experimental conditions were *forcing* the participants to establish common ground among all the members. We also pointed out situations, unlike those studied in that paper and like those studied here, in which it would be highly unlikely that the dialogue participants attempt to negotiate common ground among all the members, while at the same time there wouldn't really be any 'overhearing' in-

volved. The *SPs* reach as high a level of grounding as the dyad actively engaged in talk (speaker and addressee), without any direct exchange of evidence of understanding, between the *SP* and the active dyad.

In this chapter, we provide empirical evidence for this claim. Namely that, at least in the dialogues studied here, the common ground is, in fact, equally accessible to *SPs* and *DAs*, while the active parties themselves, assume that despite the *SP's* lack of overt feedback, she is understanding what is being said. We argue that this is evidence for *collective states of understanding* that are not reducible to their component dyadic interactions.

It appears that in modelling multi-party dialogue in such cases, we need to account for the possibility that one participant can *stand proxy*, in terms of grounding, for another (ratified) participant. We will take up this latter issue in detail in the next chapter, where we explore experimentally, the mutual responsibilities between *DAs* and *SPs* that ‘Grounding by Proxy’ as such, amounts to.

4.2 Method

Before describing the analysis in more detail we first introduce the corpus used.

4.2.1 The AMI Corpus

The AMI Meeting Corpus (Carletta, 2006) forming the basis for this analysis, is a multi-modal (video, audio and text) set of 100 hours of recorded meetings. These consist of a set of naturally occurring and a set of scenario-based meetings, the former being the focus of this investigation. AMI is fully encoded in XML. The software provides various means for browsing and displaying the different parts of the corpus, including video, audio and dialogue transcripts. The video part of the corpus consists of close-up cameras for each dialogue participant in addition to two other cameras providing views from overhead and the corner of the room. The close-up cameras provide the means to determine states of mutual gaze among the participants. As for audio, we have used only the headset-mix wave files provided for each meeting.

In this chapter, 10 of the naturally occurring meetings - a total of 10 hours of conversation - have been analysed.

4.2.2 Side Participants to Strips of Dyadic Talk

For reasons which will become clear, in order to make claims about a speaker's assumptions regarding *SP* understanding, we extracted all strips of dyadic talk from each meeting. These are segments during which there is no explicit feedback (except 'continued attention') from participants other than speaker and addressee. This provides identifiable *SPs* and *DAs* for each dyadic segment. Based on the turn taking model in (Sacks et al., 1974) (see section 3.7.1), these dyadic segments of talk end in one of two ways:

1. **Self-selected side participant (*SP*):** a *SP* wins the floor by exploiting a gap in the dyadic talk, or she interrupts the talk mid-utterance.
2. **Nominated by Last Speaker (*LS*):** Last speaker hands the floor over to a *SP*, by directly addressing her.

It is in general a current speaker's paralinguistic behaviour (gaze and body orientation) and/or the content of her utterance (e.g. use of personal pronouns accompanied by gaze) which together determine whom she is directly addressing. When a *SP* is directly nominated (addressed) at the end of a segment, it's the same information which signals a change in the speaker's set of *DAs*. Note that the *DA* is determined through reference exclusively to the speaker's behaviour. Also we take into account that the speaker might be 'addressing' the other participant in the dyad while making a *SP* the intended recipient as when the *SP* is the 'butt' of a speaker's joke (Levinson, 1988).

4.3 Analysis of Ellipsis

At the end of a dyadic segment the participants hold certain assumptions about each other's level of understanding. One way these assumptions are made manifest is in the *context-dependent* expressions employed by the speaker. These are generally taken to cover both ellipses and anaphora.

Ellipsis is a phenomenon both in monologue and dialogue whereby single or multiple sentence constituents are omitted, presumably to enhance efficiency of communication (recall the principle of Least Collaborative Effort, section 2.3.6, which might be used to justify the use of elliptical expressions). The fully resolved content of such elliptical expressions can be recovered by reference to an earlier part of a sentence, a first conjunct or a previous utterance/sentence,

which are said to be *immediately* present in the local context. Examples of ellipsis include all Non-Sentential Utterances in dialogue (Fernandez & Ginzburg, 2002), VP-ellipsis, bare argument ellipsis, gapping, sluicing etc.

On the other hand there are anaphoric expressions which are said to ‘stand in’ for some referent, salient either in the discourse context, mentally or in the immediate physical environment of the participants. These include pronouns and definite descriptions.

Arguably, anaphora and ellipses require different notions of context for their resolution. Indeed, even if we consider any one of them alone, still neither of them would constitute a unitary concept. For example, some types of ellipsis are taken to have syntactic explanations (to do with the omission of words or constituents and hence in the remit of syntax), some semantic (their resolution requires reference to the denotations of the words omitted) and some pragmatic. As for anaphora, for example, some consider deictic (indexical) uses of pronouns as anaphoric while others disagree (see (Cann, Kempson, & Gregoromichelaki, 2009), chapters 5 and 7 for a discussion of these problems).

We should however note that none of this will really affect our arguments in this chapter, as we are appealing to anaphora and ellipses merely as an empirical measure or index of access to context, whatever notion of context that may be. The claims developed here remain only at an informal, empirical level, thus not requiring, at least at this stage, an articulation of exactly what notion of context is being appealed to.

Such context-dependent expressions are central to this analysis since they index 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 that the common ground established during the dyadic exchange is accessible to each other. More specifically, at the point when the dyadic exchange ends, we have the opportunity to compare a) the pattern of use of context-dependent expressions by the last speaker (*LS*) to the side-participant (*SP*) with b) the pattern used by the side-participant to the last speaker.

If the *LS* addresses the *SP* elliptically they are demonstrating their assumption that the *SP* grounded the antecedent referents (in case of anaphora) or utterances (in case of ellipsis) during the prior dyadic conversation. Conversely when the *SP* self-selects (interrupts), the use of such expressions demonstrates the extent to which the *SPs* can directly access the common ground or the context collaboratively built up by the other participants.

Our first level classification distinguishes four categories:

- **CD (context-dependent)**: Utterance contains Ellipsis or Anaphora (including Definite References).
- **CT (continuation of talk)**: In terms of semantic content, the utterance could intuitively be thought of as the continuation of the talk in the segment, i.e. utterance does not have a coherent meaning without the background of the dyadic talk. These do not contain any surface, explicit forms of context-dependency.
- **BC (backchannel)**: Having been ‘silent’ throughout the dyadic segment, the *SP* merely starts to backchannel again.
- **NC (new context)**: Introduction of a new context/topic.

This scheme yields the following segment types: LS_{CD} , LS_{CT} , LS_{BC} , LS_{NC} , SP_{CD} , SP_{CT} , SP_{BC} , SP_{NC} .

For a second, more fine-grained level of analysis that takes the kind of ellipsis into account we needed to further decompose the *CD* category. In order to do this, a coding protocol for such expressions was developed.

4.4 Quantifying Ellipsis

In this section we present a coding protocol that allows naive users to annotate dialogue transcripts for the different explicit/surface forms of context dependency including anaphora and ellipsis¹. This protocol was developed in order to provide a reliable measure of the surface forms of context dependency. As noted, this is used here to quantify the mutual accessibility of context, that is, the extent to which participants assume that their contributions are accessible to each other. However, there has not been a reliable method for doing this. The ellipsis protocol presented here provides a useful step in this direction. It provides a standardised coding scheme that can quantify the extent to which speakers can directly access the constituents of each other’s turns.

¹The protocol presented here has been published in the proceedings of the SIGdial workshop: (Colman, Eshghi, & Healey, 2008)

Previous work has attempted to identify a taxonomy of Non Sentential Utterances (NSUs), e.g. (Fernandez & Ginzburg, 2002; Fernandez, Ginzburg, & Lappin, 2007; Schlangen & Lascarides, 2003). These are, as we have noted previously, fragmentary but intuitively complete utterances, exclusive to dialogue that are not sentential in their outward form.

The NSU taxonomy developed by Fernandez and Ginzburg (2002) has greatly influenced and guided our choice of ellipsis categories here. However, one issue with these previous approaches is the lack of reliability data; a statistic such as Cohen's kappa (see e.g. (Carletta, 1996)) is needed in order to demonstrate that a taxonomy or coding scheme can be reliably applied between independent coders. The approach presented here provides a reliable coding scheme that has been designed to be used without any specific prior knowledge of linguistics being required.

In order to simplify coding and increase reliability, some of the categories suggested by Fernandez and Ginzburg (2002) have been collapsed into broader ones. It should be pointed out that we are *not aiming to produce an accurate or definitive analysis of ellipsis*. The protocol is rather the product of contending with the compromise between robust coding categories and linguistic accuracy.

4.4.1 The Coding Protocol

As can be seen from Figures 1-4, the protocol consists of four binary branching decision trees that are applied to each contribution in an interaction. The categories presented here (in boldface) are generally ordered in terms of frequency of occurrence in order to assist the coder. Note that a contribution to dialogue may indeed contain several instances from these categories. Thus, contributions are not assigned to one category alone. Rather, coders should use the protocol iteratively on each contribution to label any part of a contribution that is elliptical, while each time ignoring the section of the contribution that was already coded. This means that the protocol has another advantage: there's no need to specify in advance what 'units' are to be coded. Rather, the units are those that the protocol itself picks out via repeated application on the same utterance.

4.4.2 Inter-rater Reliability

In order to demonstrate reliability between coders, two coders (one computer scientist, one psychologist) applied the ellipsis protocol to a sample of task oriented dialogue. This was taken from the HCRC Map Task corpus (Anderson et al, 1991); a series of dialogues in which one participant attempts to describe a route on a fictional map to another. The longest of these dialogues was

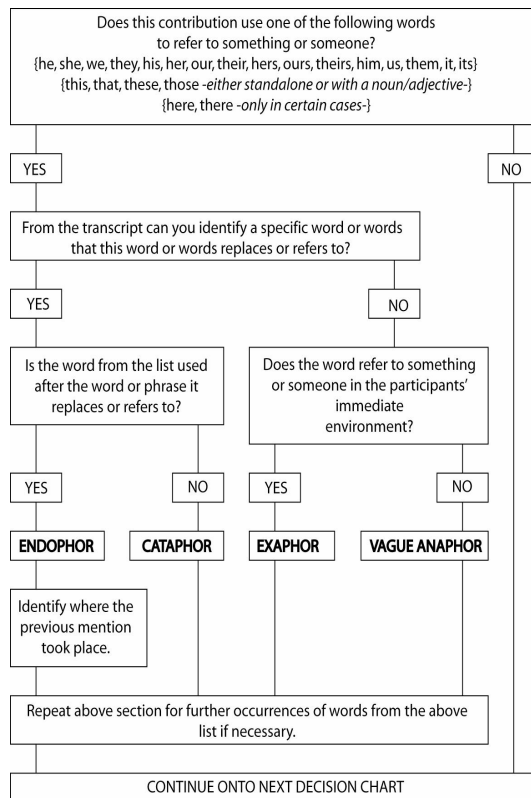


Figure 4.1: Anaphora Decision Chart

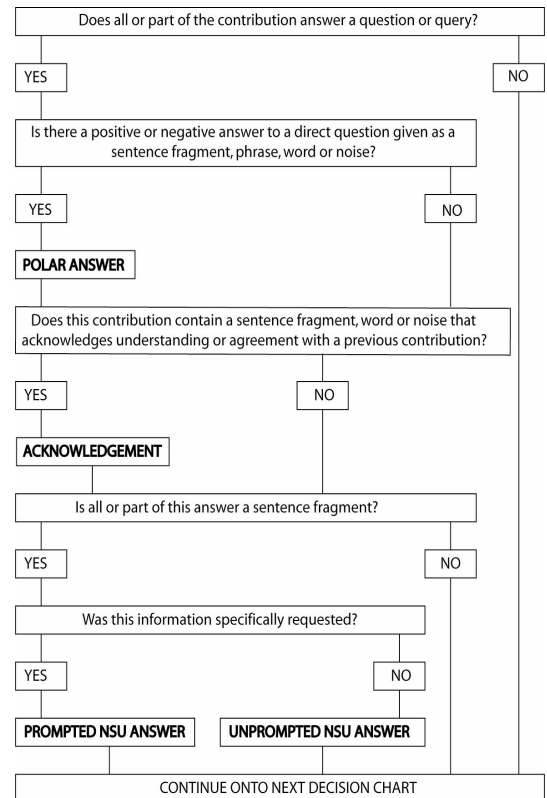


Figure 4.2: Answers Decision Chart

chosen to be coded (transcript Q1NC1) which consisted of 446 turns and 5533 words. Cohen's kappa was calculated using the procedure outlined in (Howell, 1997); see (Carletta, 1996) for a discussion of the use of kappa in dialogue coding. Kappa in this instance was .81, which shows very high reliability, even by conservative standards (Krippendorff, 1980).

4.5 Ellipsis Taxonomy

1. **Non-Sentential Utterances (NSU):** These have been coded for, using the protocol developed in the previous section. We have further collapsed the categories according to their role/function in conversation, into the following more general ones:

- **Direct Answers (DA):** Fragments used as answers to questions. Includes *Polar Answers*, *Prompted NSU Answer* and *Unprompted NSU Answer*.
- **Clarification Requests (CR):** Fragments in question form, used to request clarification or further elaboration of a previous utterance. Includes *Clarification Ellipses* and *Sluices*.

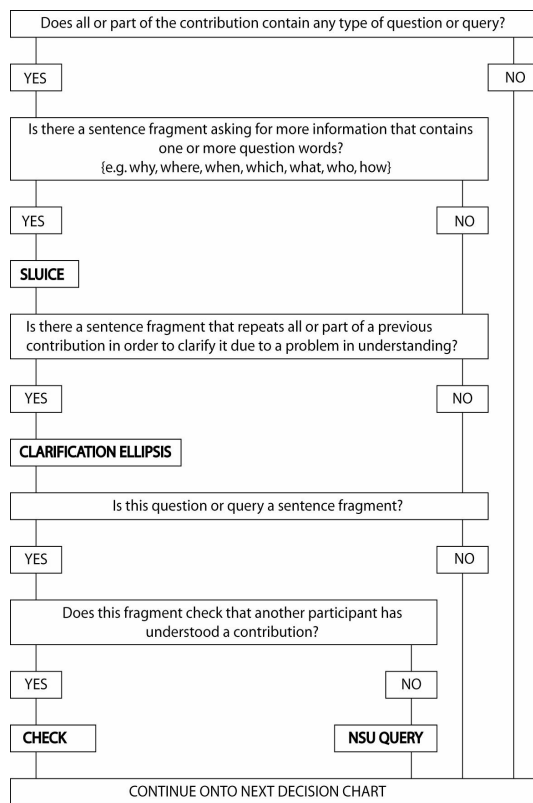


Figure 4.3: Questions Decision Chart

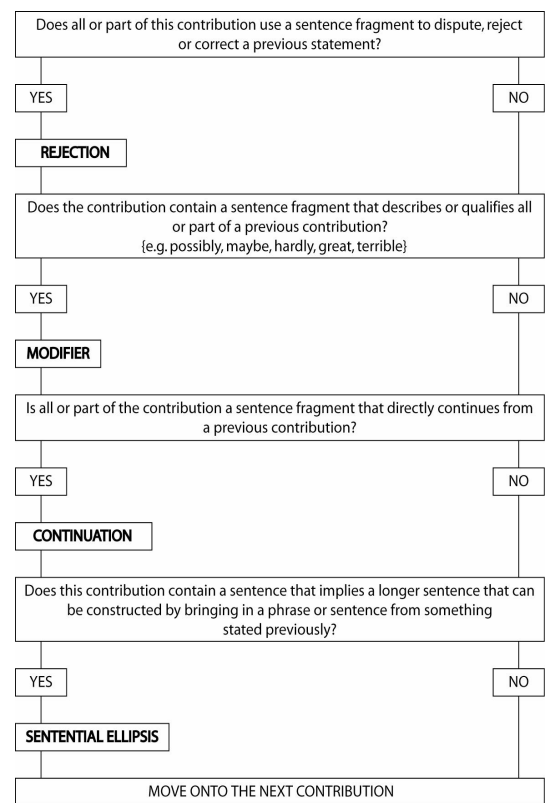


Figure 4.4: Statements Decision Chart

- **Modifiers (MOD):** In their fully resolved form, these are statements somehow modifying a previous utterance in conversation. They include *Modifiers* and *Continuations*.

2. **Sentential Ellipsis:** These are contained in utterances which are sentential, but semantically incomplete as a result of either the full omission of a syntactic constituent or its replacement by a bare auxiliary. These are captured at the very end of the protocol, that is, at the end of the Statements Decision Chart (Figure 4.4). In the case of stand-alone uses of propositional attitude verbs (know, see, believe etc.), the whole of the antecedent utterance can get elided. Note that this latter kind of ellipsis is what is commonly referred to as Sentential Ellipsis in the linguistic community. So our usage here is different and broader in that it is used in contradistinction to Non-Sentential Utterances, where the elliptical utterance is sentential in its outward form. Hence the term Sentential Ellipsis as used here includes, but is not exhausted by the usage common in the linguistic literature. Here's an example of a typical form of sentential ellipsis:

Verb Phrase (VP) Ellipsis:

A: Will you please go to the market tomorrow?

B: I already told you I will. [Resolved Content: “I already told you I will go to the market tomorrow”]

Since the protocol developed does not determine what kind of Sentential Ellipsis is being used, we have developed an ad hoc taxonomy analogous to that for NSU’s, based on the role/function of the utterance containing the ellipsis. Bear in mind that the taxonomy is being used merely to compare what *SPs* and *DAs* can ‘do’ elliptically.

- Direct Answers (DA): Utterance containing the ellipsis is an answer to a question, like the above example.
 - Request for confirmation (RC): Partly redundant, these are tag questions used to request confirmation or initiate dispute. “A: I got an A in Biology. B: Did you? A: Yes. I got the results today.”
 - Statement (ST): General category containing all statements, excluding Direct Answers.
 - Query: All elliptical questions excluding Requests for Confirmation.
3. **Anaphora (Anaph):** These were also coded according to the protocol developed in the last chapter.
4. **Definite/Nominal Reference (DR)**

To provide a baseline comparison of ellipsis types in natural dialogue we also coded 10 peoples conversations from the British National Corpus (BNC).

4.6 Results and Discussion

Table 4.1 shows dyadic segment type counts, for 10 AMI meetings.

4.6.1 Segments of type LS_{CD} : Assumptions about *SPs*

All such segments indicate that the last speaker, in producing elliptical utterances addressing a *SP*, is tacitly making the assumption that the *SP* would be able to resolve the ellipses employed, which in turn depends directly on the *SP* having grounded the antecedent utterance(s)

	<i>CD</i>	<i>CT</i>	<i>NC</i>	<i>BC</i>
<i>LS</i>	20	4	3	4
<i>SP</i>	100	33	1	0

Table 4.1: Dyadic Segment Type Counts

of the ellipsis contained within the segment, for which the *SP* did not produce any explicit feedback. Note that ‘Continued attention’ by the *SP*(s) is very frequently not monitored by any of the participants in the dyad. Eye contact is more or less exclusively maintained between the two of them. Nevertheless the *SP* is ‘expected’ by the last speaker to have grounded the antecedent utterance(s). Furthermore, none of these segments were followed by any form of Repair/Clarification by the *SP*. In all of them the *SP* seems to be coping perfectly well with the elliptical utterance, and the conversation goes on smoothly. This evidence seems to support the claim in (Branigan, 2006) that speakers have very similar assumptions about the level of grounding achieved by *SP*s compared to that by *DAs*. Nevertheless Branigan also proposes that speakers often have weaker expectations from *SP*s.

The following are excerpts from AMI, showing the different kinds of ellipsis employed by the last speakers:

Anaphoric chains: distant antecedent recognised by SP

- (4.1)
- (1) B: Yeah. But that still won’t tell you. well howmany **tangrams** are there that they’re using? Fifteen or something.
 - (2) C: Uh no, not even that. They’ve of this relevant type
 - (3) B: Uh-huh. So that’s not gonna so that’s not gonna tell you anything about **their** relative complexity. . . You still need some kinda scale for **these things**. Ca uh if you look at **em**, do you just know
 - (4) C: Mm no. [laugh] Well I don’t. I’m not .
 - (5) B: No. I wouldn’t either. What about him? I if Mister Geometry. I mean, you know. Can you tell just by looking at **these** how hard people find them?
 - (6) A: No, I’m afraid not. I wouldn’t know.

In the above excerpt, also note how similar C’s (the *DA*) last utterance is to A’s (the *SP*): VP ellipsis in C’s versus whole sentence ellipsis in A’s utterance.

The whole segment as antecedent

- | | | |
|-------|-----|--|
| (4.2) | (1) | B: [. . . 6 utterances so far exchanged between B and C] Data I think we should keep in. |
| | (2) | C: OK. [laugh] |
| | (3) | B: Because it's would be the same as feature. |
| | (4) | B: Or spec spectrum. I think data's the same as spectrum . . . |
| | (5) | C: I do I still don't think that goes in. But . |
| | (6) | C: yeah, I still don't like it. But |
| | (7) | B: Final view, Bob? |
| | (8) | A: I don't have passionate feelings. |

Here, B's last utterance explicitly addressing A, is highly elliptical with no particular utterance as antecedent, i.e. the resolved content of the utterance depends on the whole segment between B and C. B expects A (Bob) to have grasped the issue under discussion. One would expect A here to initiate clarification if he really didn't know what B was asking.

We think that the speaker's assumptions about *SPs* are, amongst other things, strongly mediated by the speaker's prior beliefs (before the conversation) about the *SP* and his relevant knowledge. In the meeting from which the above was extracted, A is a supervisor with whom the rest of the participants check their results as they go along. So, firstly if cooperative, he should be 'paying attention' to the dyadic interactions in which he is not directly involved (most of the meeting). Secondly, the rest of the participants believe to begin with, that he would understand such technical issues under discussion. So perhaps it is some notion of the well known 'lab coat effect' that could justify such high expectations.

4.6.2 Segments of type *SP_{CD}*: Side participant access to 'communal' common ground

In this section we will argue that *SPs* have the same kind of access through the same techniques, to the 'communal' common ground, as the participants directly involved in collaboratively securing it (speaker and addressee). These segments, which comprise the largest class in this analysis, end when a *SP* interjects producing an elliptical and hence context-dependent utterance. Again here, the antecedents of the ellipses, lie within the dyadic segments.

Table 4.2 below shows the ellipses identified in these *SP* utterances. They have been classified according to the taxonomy described in section 4.5. In order to assess whether there is a difference between the use of ellipsis types by *SPs* and the baseline - typical frequency of use independent of both the number of participants in the conversation and the status of the participant

	Non-Sentential			Sentential				Other		Tot
	DA(NSU)	CR	MOD	ST	DA	Query	RC	DR	Anaph	
AMI	14	9	12	10	0	4	6	11	58	124
Baseline (BNC)	154	73	46	46	21	12	20	not coded	303	675

Table 4.2: Ellipses employed by *SPs* terminating *SP_{CD}* segments compared to the baseline (BNC)

upon employing the ellipsis - used in ordinary dialogue, we compared the frequency of ellipses of each type with that found in the BNC. Taking into consideration all categories in Table 4.2 (merging Sentential and non-Sentential DAs and ignoring DR since it wasn't coded for in the BNC) there is a reliable difference ($Chi^2_6 = 14.6$, $p = 0.02$). However, as Table 2 indicates the main difference is in the relative frequency of direct answers which account for 26% of instances in the BNC but only 12% of instances in AMI. If this category is ignored we find no reliable difference between *SPs* and the baseline ($Chi^2_5 = 4.33$, $p = 0.50$).

The difference in frequency of use of direct answers is essentially an artefact of our coding scheme. As noted above the *SP* ellipses are those where they have nominated themselves as next speaker by interjecting. Consequently, direct answers by *SPs* to questions - which are by definition not directed at *SPs* - are much rarer. Subject to this caveat, we can conclude that the pattern of use of different ellipsis types by *SPs* is not in fact distinguishable from the pattern of use typical of participants in ordinary dialogue.

What now follows is a discussion over a set of examples from AMI comparing the kinds of access to context through various elliptical phenomena, possible by *SPs* to those by *DAs*.

Anaphora with distant whole utterance as antecedent

- (4.3)
- (1) B: Um so this person didn't ha um the obviously didn't know about capitalisation. So just about every utterance needs to be capitalised and and needs the end punctuation.
 - (2) D: Mm-hmm
 - (3) D: You know, **when you get like um someone's talking and there's they sort of pause in the middle of a sentence that's long enough for it to put a break in,**
 - (4) B: Yeah
 - (5) D: **but they're actually sort of carrying on the sentence, do you have to capitalise each time you transcribe a bit of it's mid**
 - (6) B: Um, no no no.
 - (7) D: No no no no. Yeah.
 - (8) B: Whatever um makes sense to you.
 - (9) D: Okay.
 - (10) B: Um [cough] but no, it it can continue into the next segment and that's perfectly fine. the hyphen and then.
 - (11) C: Yeah.
 - (12) C: I think **that's** actually the only case where you don't
 - (13) C: or where you're not supposed to capitalise, right?

Utterances 1 to 12 above form a segment of type SP_{CD} which is terminated by *C*. The anaphora 'that' in 13 can only be resolved with 3 as an antecedent. An issue is here raised initially by *B* to which *D* responds by asking a question (utterance 3). All the way down to utterance 12 the question is under discussion exclusively between *B* and *D*. *C* then produces utterance 13 which can intuitively be thought of as a negative answer to or a rejection of *D*'s initial question. In other words it could have been produced by *B* (the *DA*) adjacently to the initial question. Note here that *C* has had to re-raise the context in order to make her contribution. I.e. a simple "No" (a Negative Polar Answer) like *B*'s initial response, or even the less elliptical "I think that's actually the only case where you don't.", would most probably be infelicitous (the other party would be very likely to initiate clarification). But this seems to be the effect of antecedent distance alone, since all of the NSU classes are possible by *SP*s at the end of the segments in question, but clearly not at such high antecedent distances.

Non-Sentential Utterances (NSU)

Factual Modifier

- (4.4)
- (1) C: When I did my masters um I took uh SP1 and SP2 with Simon King.
 - (2) A: You survived SP1 and SP2.[laugh]
 - (3) C: Yes. And actually I've done quite well in SP1, I've done it a bit worse in SP2 because it was a l a lot more challenging.
 - (4) A: We have two new teachers for SP2.
 - (5) B: **Too many.** [laugh]

The excerpt above shows an instance of a *SP* Factual Modifier (boldfaced in the excerpt) produced by *B* adjacently to its antecedent. The same utterance “Too many.” by the *DA* (*C* here) would have been perfectly felicitous (implies in this case, equivalence of *SP* and *DA* access to context).

Among the NSU classes in (Fernandez & Ginzburg, 2002), Clarification Ellipsis (CE) has a special status, since it is known to be a common technique used in dyadic dialogue to ground utterances which weren't sufficiently understood by the recipient. There were very few CEs identified in this analysis. However, we do know that *SPs* can and do in fact initiate elliptical clarification, by exploiting gaps in dyadic talk:

Clarification Ellipsis (CE)

- (4.5)
- (1) C: What does cutest spelling mean?
 - (2) B: oh, she spelled cutest um with an I_.
 - (3) C: oh, okay.
 - (4) B: so that that's just something I pointed out.
 - (5) D: oh yeah.
 - (6) A: **Cutest?** [Gazing at D. Direct Addressee is D here.]
 - (7) D: E_S_T_
 - (8) A: Thank you. [laugh]

D and A above are both *SPs* to the dyadic segment between B and C. The CE produced by A is interestingly addressed at D who is also a *SP* to what's being clarified, which shows that in multi-party dialogue all the ratified participants have responsibilities towards one another.

This example also indicates clearly that there can be varying levels of understanding among the *SPs* themselves. However, note that we are not claiming by any means that in multi-party situations, the participants always reach a collective state of understanding. The claim is rather that such collective states do exist, and that they're often assumed by the parties involved.

Furthermore, it's interesting to see that had it been B (the *DA* of the antecedent utterance)

who didn't understand, she would have produced the CE locally (as opposed to at a distance of 5 here) as is generally expected in dyadic dialogue. This example has important implications for models of dialogue context, discussed briefly in section 4.6.4 below, and taken up again in the last chapter of this thesis.

VP ellipsis

The taxonomy described in section 4.5 has been used to classify these.

local VP ellipsis by SP

- (4.6)
- | | |
|-----|---|
| (1) | B: [5 utterances exchanged between B and C so far in the segment] but I I I do know the type of scenario you're describing. I just it's just hard to answer that without hearing something. Mm. |
| (2) | C: Mm-hmm. <i>The um should be capitalised.</i> |
| (3) | B: yeah, <i>they</i> should all. I stopped marking them , 'cause there are just too many. |
| (4) | C: yeah |
| (5) | A: Should it? 'Cause the loose uh is continuing from one sentence isn't it? |

Note also the chain of anaphora referring to the 'um' (indicated in italics), and how it carries on across to the *SP*'s (A's) utterance. This phenomenon is very frequent in *SP* utterances terminating *SP_{CD}* segments.

4.6.3 Segments of type *SP_{NC}*: Implications for our claims

The analysis indicates that the introduction of new contexts/topics by *SP*s interrupting a dyadic segment, is extremely unlikely. Consequently if a *SP* is to interrupt, she has to 'stick to the topic' already under discussion in the segment. This further supports our claims, in that even if a *SP* is not using ellipsis as direct access to the 'communal' common ground, she makes use of the information in there, to produce a relevant utterance. An utterance thus produced, semantically depends on and is incoherent without the background of the 'shared knowledge' established between the speaker and addressee in the dyad.

4.6.4 Context locality/saliency

Recall that in section 3.11, we pointed out that the transformations proposed in (Ginzburg & Fernandez, 2005), to handle multi-party dialogues, are based on 2 empirical benchmarks. One of these was the adjacency of grounding and acceptance acts to their first pair-parts. We argued that this benchmark bears closer inspection. Among the examples given above, we can find at least 2 which contradict this benchmark, and hence render the corresponding transformations unfit to handle such cases:

- Grounding/Clarification at Distance: The Clarification Ellipsis in the example given above (Excerpt 4.5), has an antecedent distance of 5, that is, it is non-adjacent to its antecedent.
- Discussion/Rejection at Distance: Excerpt 4.3 is a case of non-adjacent rejection at a distance of 10 turns.

We will not discuss the exact nature of the problems these examples create for Ginzburg's model here, but will come back to them in the last chapter of this thesis, where a solution will be sketched.

For now, note that in excerpt 4.5 A's CE is, in effect, just as local to C's question utterance as B's adjacent response. What this and similar examples make clear is that locality or saliency of an antecedent cannot be a simple function of time or turns elapsed. An alternative notion of saliency/locality is developed in chapter 6, where this same example is used to motivate our experimental hypothesis.

4.7 Conclusion

The evidence from this analysis shows that, with respect to common ground, side-participants in the AMI corpus do not appear to be different in any substantive respect from direct-addressees. Speakers assume that *SPs* reach the same level of understanding as the addressees. Additionally, *SPs* were shown to use elliptical techniques to access the shared-context, in generally the same way as *DAs* do. All things being equal, this is strong evidence for collective states of understanding that could not be predicted from considering the component dyadic interactions alone since, *prima facie*, the *SPs* don't ground to the same level and don't go through the same grounding cycle as *DAs* with the speaker.

As mentioned in the introduction, the corpus evidence presented here indicates that direct addressees can act as *proxy* for side-participants in providing evidence of understanding or more

broadly in carrying the conversation forward. *Grounding by proxy*, as such, will be explored experimentally in the next chapter, where we also present conditions under which it is expected to hold. As we will see, it will turn out that this entails a set of mutual responsibilities between direct addressees and side-participants.

Chapter 5

Parties in Conversation: Grounding by Proxy

5.1 Introduction

As sketched in chapter 2, in dyadic dialogue, conversational responsibilities take the form of moment by moment collaboration and exchange of evidence of understanding between speaker and addressee. The systematics of the grounding process as such, are relatively well understood and spelt out in considerable detail by the work of Clark and colleagues.

In multi-party dialogues on the other hand, the grounding process is more complex and understood less well. The participants may, at any moment, occupy different roles depending on the relation that they stand in, to an utterance: speaker vs. side-participant vs. direct addressee vs. overhearer. We pointed out, in section 3.2.9, that such differences are consequential for the participants' (expected) behaviour in conversation, as incumbents of those roles (see e.g. Goodwin (1984); Levinson (1988)). We also discussed in section, 3.4, how such differences in role lead to systematically different patterns of common ground being established between the participants (Wilkes-Gibbs & Clark, 1992; Schober & Clark, 1989).

The studies alluded to above, of patterns of grounding in multi-party dialogues, all relate to the difference between ratified participants and overhearers in conversations. However, being a side-participant is clearly also different from being an addressee of an utterance, in that "some answer is ... anticipated from them [addressees], more so than from other ratified participants [side-participants]" (Goffman, 1981), meaning that side-participants have 'weaker' responsibilities as to providing moment by moment evidence of understanding for the speaker.

Nevertheless, in the previous chapter, we showed that despite this imbalance between the

levels of feed-back provided by side participants and direct addressees, side participants may still reach a level of understanding comparable to direct addressees. In other words, withholding of negative feedback from side participants is taken as positive evidence of understanding. The side-participants were being treated as equivalent to direct addressees in the level of understanding they were assumed to have reached. We concluded that such dialogues lead to *collective states of understanding* that are not reducible to their component dyadic interactions.

We should point out here that the results from the previous chapter are in conflict with those of Healey and Mills (2006) sketched in section 3.8. They showed that differences in status even amongst the ratified participants (side-participant vs. speaker and addressee), can be consequential for the levels of semantic coordination achieved by them. Healey and Mills (2006) show that within one and the same local conversational context, people distinguish between primary (those actively contributing) and peripheral participants (those who aren't actively engaged in the interaction but are nevertheless ratified): the same question receives systematically different answers depending on whether the apparent source of the question was a peripheral or a primary participant. They suggest that this is due to the opportunities that interaction creates for the participants to address differences in interpretation, and, as a consequence, those passive, peripheral participants who do not actively contribute - that is to say they do not actively provide feedback - are treated differently. In other words, it is the imbalance between participants' level of participation which is leading to differences in levels of semantic coordination.

But note that the dialogues from the AMI corpus used in the previous chapter were of a special type: they were meetings in which all the participants were expected in advance to sustain a more or less balanced level of participation or involvement in the dialogues. This is the crucial difference between the Healey and Mills (2006) peripheral participants and the AMI side-participants. But what exactly is the difference here in terms of grounding behaviour? What 'extra' responsibilities do these 'active' side participants have? - that is sufficient for them to be treated as 'present' in the current conversational context and hence the conversation moving forward and common ground accumulating for them in the same way as it does for the others? Below, we suggest one possible answer.

5.1.1 Grounding by proxy

So how should these collective states be modelled? I.e. what grounding responsibilities, if any, does being an 'active' side-participant to an utterance in such dialogues entail? Recall the

Principle of Responsibility:

In a conversation, the parties to it are **each** responsible for keeping track of what is said, and for **enabling the other parties** to keep track of what is said. (Clark & Schaefer, 1992)

We pointed out, in section 3.5, how this principle, is really just a normative description of how ratified participants should behave with respect to grounding what is said, and that in multi-party dialogues, particularly as regards side-participants, it does not provide an account of how such responsibilities are to be fulfilled at a local level. In this chapter the aim is to operationalise this principle and, as we will see, Schegloff's notion of 'parties', in conversations of the type studied in the previous chapter.

Results from the previous chapter suggest that we may need to account for the possibility that a currently engaged/active (ratified) participant (the current speaker or addressee) may *stand proxy* for side-participants of the moment, with respect to grounding what is said or more broadly in carrying the conversation forward, while these side-participants are, in withholding of repair/rejection, assumed to have understood/accepted what was said.

As such, *Grounding by Proxy*, if operative, does indeed entail a set of responsibilities, borne by the side-participant of the moment, with respect to something uttered by the speaker or addressee (as also implicit in the principle of responsibility above). Namely, if a speaker/addressee stands proxy for a side-participant in carrying the conversation forward, then if for any reason (s)he fails to carry the conversation forward, thus not satisfying the requirements imposed by the current context of the conversation (e.g. providing the second pair part implicated by a first already uttered), then the side-participant should compensate by interjecting in order to satisfy those requirements. Such failure on the part of the current speaker, can be manifest as signs of mis/non-understanding, hedgings, pauses, inconsistencies and falsities, as seen from the side-participant's point of view. Such speaker/addressee failures may also get manifest, as explicit elicitations of help by them from the side-participant.

For a simple, initial example, consider the dialogue below:

- (5.1) | (1) Alice: What time is it **guys**?
 | (2) Alan: errrr . . [looking around to find a clock]
 | (3) Brian: Its five thirty.
 | (4) Alice: Oh Thanks.

Alan's failure above to address the obligation imposed by Alice's question - namely to answer - promptly, is followed by Brian's fulfilment of it. We will explore below the different configurations of the participation framework in which this may occur.

5.1.2 When is grounding by proxy expected to hold?

Surely, such mutual responsibilities, are not always operative. In particular, there are dialogues - or stretches of talk or topics therein - during which one or more of the ratified participants, do not have any incentive of their own to move the conversation forward or contribute. They may well be indifferent to what's being said at the moment. They may occupy a totally passive role (recall e.g. the peripheral participants in Healey and Mills (2006)).

Instead, we'd expect the responsibilities entailed by grounding by proxy to be in effect, in case the side participant of the moment has similar degrees of incentive to understand/ground/respond, as the person who already holds the floor. So we might say that grounding by proxy holds when two or more participants, form a single 'party', or 'collectivity' with respect to the current, local communication task at hand. Let's explore in more detail what consequences such 'party' formations might have for the organisation of subsequent talk/grounding.

5.1.3 Parties in conversation

Schegloff (1995) introduced the notion of a *Party*¹ in multi-party dialogues, as comprised of more than one person. He suggests that the simplest systematics of turn-taking proposed by Sacks et al. (1974), summarised in section 3.7.1, operates not in the first instance between individuals but between parties:

But on some occasions, or for a particular phase or topic or sequence [locally] within some occasion of talk-in-interaction, the aggregate of persons who are as Erving Goffman called them ratified participants are organised into parties ... by virtue of interaction specific contingencies and conduct. (Schegloff, 1995)

The idea is that certain topics or phases during the course of the conversation, i.e. the current conversational context, may project or make relevant, the coalition of a group of individuals. In other words, the context can be such that a certain subset of the participants are in alignment

¹Note that in other places, including work by Clark and colleagues and the previous chapters of this thesis, this word has been used exclusively to designate a single individual. But we will hereafter use this word to refer to a subgroup of the ratified participants (possibly a single individual), a notion which is central to this thesis and explicated further in this chapter.

against the background of that context, e.g. when two couples are discussing plans to meet at some later time, in which case the discussion is taking place between two parties (in this case two couples) rather than 4 individuals. So the members of the party are equally expected to have something to say in the current context. Empirical support for this, comes from the observation that often all of the simultaneous talk in conversation is between members of one party, as the members compete for the floor to make a contribution to carry the conversation forward (Schegloff, 1995). They are as responsible for satisfying the constraints imposed by the current context as the other members of that party. In such circumstances, we may say that one person in such a coalition may speak on behalf of the rest of the members. The competition for the floor in such cases indicates that if one of them somehow fails to address the issue at hand, then the others would compensate, so that the party is responsible as a whole to address that issue; hence *grounding by proxy*.

When parties are contextually relevant, an utterance from one party member, as already noted, could, in withholding of negative feedback from other party members, be understood to stand proxy for what they would have said. One can, in this manner, ‘speak on behalf’ of a co-present group. Rom Harre’s (1999) concept of ‘positioning’ in intergroup relations is particularly relevant here. Tan and Moghaddam (1999) use this concept to make a distinction between actions which are ‘positioned’ as belonging to an individual and those that are positioned as belonging to a collective:

In such instances, and particularly when one’s status as having ‘the right to speak on behalf of’ is formally recognised, a person’s utterances are likely to be taken as a group commitment, rather than as an individual undertaking. The illocutionary effect of what is said extends to the group on whose behalf the representative speaks.
(Tan & Moghaddam, 1999)

Thus, the commitments brought about by something uttered by the speaker of the moment, could, in withholding of dissent from the (possibly side-participant) party co-members, affect the party as a whole. This chapter is an experimental investigation of the situations in which this may occur, and of the consequences this may have, for the organisation and structuring of talk, grounding and the interactional patterns within a group, as implied by the notion of grounding by proxy.

5.1.4 Projection of parties and Corresponding Configurations of the Participation Framework

Lerner (1993) expanded on Schegloff's notion of parties. He distinguished four ways in which parties may be projected or made relevant in interaction, depending on whether the speaker or the recipients are members of the prospective party:

1. Collective Addressing (e.g. 'you guys'): Speaking to/addressing more than one person as a party.
2. Collective Speaking (e.g. 'we'): Speaking for and as a party.
3. Conferring among the members: One member speaking to another member of a party, concerning the party.
4. Speaking by someone outside a party about the party, addressed to another person outside the party (e.g. 'they').

Since we are, as noted, here mainly concerned with the consequences of party formations for the organisation and structuring of talk and grounding, we can assume that it is already established and relevant. This may have been via the above or something that extends beyond the specific talk taking place or an 'extra interactional tie' as Schegloff (1995) would say (e.g. a couple, where the couple as a whole is contextually relevant). In any case, the above serve as guidelines as to the different configurations of the participation framework, determined by an utterance with respect to which certain obligations need to be fulfilled collectively by the party as a whole.

In case 1, the party is being addressed by an outsider with a sequence initiating utterance, or a first pair part (e.g. a question). Here, the party act together to satisfy the obligations imposed by this first pair part. The producer of the second pair part thus stands proxy for the rest in satisfying the obligations imposed by the first pair part. So this case relates to the mutual grounding responsibilities holding between non-speakers of the moment with respect to something uttered by the speaker. Note that in this case, a collective address (implicit or explicit) in the sequence initiating first pair-part is not necessary. The speaker of the sequence initiating utterance may be addressing a single member or all the members of a party, while the party remains relevant as a whole. This gives rise to the following sub-cases:

- **1a:** The sequence initiating utterance is addressed to the party as a whole: in this case a direct addressee self-selects (in terms of turn-taking) and stands proxy for the rest of the *addressees* in responding. E.g. dialogue 5.1 above. Note that depending on the content of the first pair part, it may well be that a response is required from every member.
- **1b:** The sequence initiating utterance is addressed to a single member of a party: in this case the direct addressee stands proxy for the rest - *side participants* in this case - in responding.

In case 2 above, one member of the party speaks on behalf of other members of the party. This can be achieved by the use of the plural first-person pronoun ‘we’ or tacitly via semantic content. To such an utterance, the rest of the members are side-participants. On this occasion, this one member, stands proxy for the rest in *speaking*.

In case 4, analogously to case 1, talk is expected from the party as a whole, where the members are at that point all indirect recipients, i.e. a side-participant expected to respond, e.g. when the side-participant is the butt of a speaker’s joke.

So we are dealing with 4 configurations of the participation framework, with respect to the utterance which gives rise to the local conversational obligations to be collectively fulfilled. These are 1a, 1b, 2 and 4. The mutual obligations holding among the members of the party are between: Addressee/Addressee, Side-participant/Addressee, Side-participant/Speaker and Side-participant/Side-participant respectively.

If grounding by proxy, as sketched in the previous section is in effect, then in all of the four cases above, the co-members, are responsible to ‘assist’ or ‘compensate’ in carrying the conversation forward, if anything goes wrong. Such assistance can either be elicited by the speaker (a member of the party) holding the floor, or a co-member can interject. The former certainly gives rise to what Lerner (1993) has called ‘Conferring’ (case 3 above). The latter can also give rise to conferring among the members but need not (recall dialogue 5.1 above where a second addressee interjects to answer the question directly without any form of conference with the co-member). So, case 3 above, apart from being one way in which parties are projected, is one of the important ways in which grounding by proxy could operate.

But, what experimental evidence is there that grounding by proxy as described above operates among the members of a party? Or whether such party formations do not fail to have an impact on the grounding behaviour of its members? In this chapter we will attempt to provide causal

evidence for the practical reality of parties, expressed as the mutual responsibilities entailed by our notion of grounding by proxy, that we will show, hold between the members.

Before we can do that however, we need to have some idea about what sorts of contribution the assistance/compensation on the part of a co-member described above, amounts to. Therefore, some instances of what we expect these to be, are now in point. At least the last 4 cases below are instances of ‘conferring’ (case 3 above). The first may or may not be:

- **Collaborative Completions or Split Utterances:** These are anticipatory completions of another’s turn/sentence. (cf. Lerner (1991)); in our case by a co-member of a party. Here, we have 2 sub-cases:

- Both parts of the completed turn may be addressed to an outsider, pertaining specifically to case 2 above, like the following example:

- (5.2) | (1) Dad: we take care of all of our business but any time we go contrary to
any of the rules, here errm [pause]
(2) Carl: ih [we] take a unanimous vote

Adapted from (Lerner, 1993)

- The first part addressed to an outsider, but the second addressed back at the producer of the first. Here, we would expect the second part to be a check question:

- (5.3) | (1) A: So we both had errrm [addressed to an outsider]
(2) B: curry? [addressed to A]
(3) A: oh yeah curry
(4) C: ooo I love curry

- **Turn Assignments:** These are explicit speaker selections, which here, are either addressed by the current speaker at a co-member, thereby eliciting help, or addressed by the co-member at the speaker, thereby reminding them that they need to say more. E.g. ‘John?’
- **Other-initiated repair:** These are repair initiations addressed by the co-member at the current speaker, in case they don’t understand what is being said.
- **Other-repair:** These are repairs done by the co-member on the current speaker’s turn. This may in fact be one environment where other-repairs may not be dispreferred or at least less so than in dyadic dialogue, as the current speaker is speaking also on behalf of

the co-member and hence has the right to modify it directly. But this certainly needs a separate analysis to be confirmed.

- **General discussion/rejection:** Similar to other-repairs these are addressed to the current speaker by a co-member, thereby initiating discussion/disagreement about what the speaker is telling a third person outside the party.

5.1.5 Ramifications of Grounding by Proxy for Participant Role Distinctions

Note that if conversational responsibilities are to be indexed to participant status at the level of the utterance (which is roughly the idea entertained by Clark and Schaefer (1992)), then grounding by proxy if operative at any level, would mean that we need to distinguish between the status of those who fall outside a party and those inside. In other words, we need to distinguish between currently ‘active’ vs.. ‘passive’ side participants and hence recognise degrees of peripherality. This is because a person outside a party has a different stance (as Goffman would say) to the current utterance. The (grounding) behaviour expected from them would be different. Moreover and crucially for models of multi-party dialogue context, this should determine the assumptions made by the currently active participants (speaker/addressee) about the level of understanding achieved by the side-participants; viz. whereas withholding of repair/rejection from an ‘active’/‘concerned’ side-participant would be taken by others as positive evidence of understanding/acceptance, for passive bystanders (e.g. the Healey and Mills (2006) peripheral participants) this isn’t necessarily so.

In this chapter, the aim is to test the idea of grounding by proxy experimentally, in order to see whether such mutual responsibilities are actually in effect between co-members of a party. We therefore attempt to create experimentally a situation where a party is projected via the goals of the task. As will become apparent shortly, our experimental manipulations are such that the members of the party will be faced with situations corresponding at least to cases 1a and 1b above. Hence our experiment counts as a test of whether non-speakers of the moment - who are members of one party - are responsible as a whole to satisfy the constraints/obligations imposed by the speaker’s first pair-part; i.e. whether one non-speaker can stand proxy for another, in *responding* to or grounding something uttered by the speaker.

Note that the local, utterance-level compensatory devices described above, aren’t what we’re solely concerned with. It may well be that failures by one party member to address a contextual obligation at hand leads to a global effect, at the conversational level, in other co-members’

behaviours. For example, such failures by *A* might reduce other co-members' confidence in *A*'s ability to address the issues under discussion, thus causing a change in the overall balance of these others' participation in the dialogue. As such, this analysis is not limited to the mutual responsibilities holding among the party members at a local level, but more broadly that of the effect of one member's difficulty to ground, on the conversational context in general and accordingly on the subsequent behaviour of others.

5.2 Method

5.2.1 Task and Materials

The experiment reported here, uses an on-line 3-person variant of the Maze Task used in studies of semantic coordination (e.g. (Garrod & Anderson, 1987; Healey & Mills, 2006)).

The Maze Task

The original task involves a pair of participants who go through a sequence of 12 mazes, one pair of mazes at a time (one maze for each participant). The mazes are based on a 7x7 grid of nodes/locations, but with different configurations of these nodes for each instance of the maze (see e.g. Figure 5.3). For each maze, some points on the grid are instantiated as nodes while some remain empty. Each node is potentially connected to an adjacent node via a path which in turn may be blocked by a gate. Some nodes are coloured in grey which designates them as switches. On each maze, there is a unique finish point marked by a red X. A cursor indicates to each participant where (s)he currently is on the maze. They can move around on the maze using the arrow keys. The mazes in each pair, received by the two participants at each stage, are identical in shape, i.e. they have the same configuration of nodes. However, the positions of switches, gates and the finish point are different on the two mazes. Furthermore, nothing on a participant's maze indicates where the other participant is. Each pair of mazes is solved when both participants reach their finish points. In order to get there, they need to open the gates at various stages to pass through. The gates on participant 1's maze will open if participant 2 goes on participant 1's switch and vice versa. But recall that the positions of switches are different on the two mazes. So participant 1 needs to communicate the position of his switches to participant 2 and vice versa. This creates the recurrent need for the participants to collaborate and communicate maze locations to one another in order to navigate through the mazes.

Our variant

The experiments reported here use an online 3-person variant of the maze task just described. Two of the participants are randomly assigned to *the Duo role*. This means that they are in the same ‘team’, as the mazes that they receive at each stage are absolutely identical in all respects (see Figures 5.1, 5.2); viz. they have identical shapes, the same switch, gate and finish positions. The third participant is assigned to *the Solo role*. The Solo receives the same maze - in terms of shape - but with different switch, gate and finish positions (see Figure 5.3). The gates on the Duos’ mazes will all open if the Solo goes on any of their switches. The gates on the Solo’s maze will open if the Duo are **both** on one of his switches at the same time. The task thus ensures that the Duo must co-ordinate their actions with each other as well as with the Solo to be successful. Each maze is solved if all three reach their finish points. Communication between the participants is via the DiET Toolkit described below.

5.2.2 The DiET Experimental Toolkit

This is a custom built Java application which allows 2 or more participants to communicate via a shared chat window similar to proprietary instant messaging systems². The display is split into an upper window, a status bar and a lower window (see e.g. Figure 5.4 below). The upper window displays the ongoing conversation, and the lower window is used for typing. All keys pressed are recorded and relayed to the server where they are time-stamped and stored. The status bar, a prominent single line of text that is controlled by the server indicates whether other participants are currently typing.

Everything that the participants say to each other, passes through the DiET server. The server stores all the relevant details about the utterances sent, including timing and structural information (the turns can be parsed in real-time on the server). The server provides fine-grained, character by character control over what each participant can see on their chat window. Different experimental manipulations or interventions can be programmed, as appropriate, into the server. These range from changing the apparent origin or the content of a real turn sent by a participant, to sending fake probe questions to a participant which appear to originate from another and capturing the participant’s response. All of this can be carried out without being noticed by the participants (see Healey, Purver, King, Ginzburg, and Mills (2003) for a detailed description of this technique). Moreover, like the maze game described above, different experimental tasks can

²<http://www.dcs.qmul.ac.uk/resrarch/imc/diet/index.php>

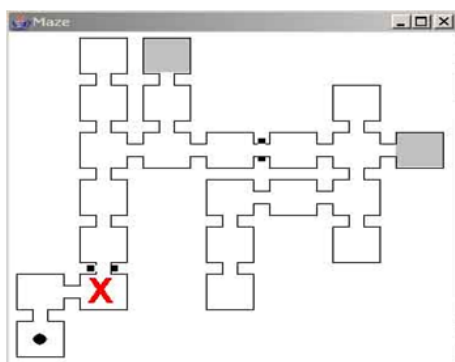


Figure 5.1: Duo 1 Maze

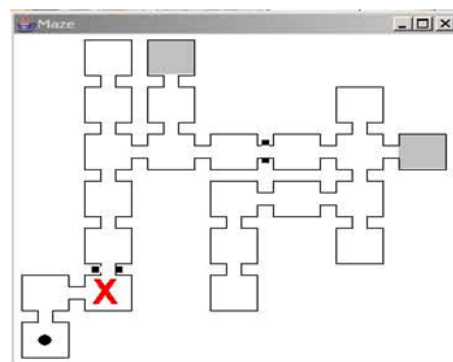


Figure 5.2: Duo 2 Maze

be programmed into the server. This allows task-level events to be detected and used to trigger interventions.

The DiET toolkit has been used successfully in a number of experimental studies of semantic coordination in dialogue including Healey and Mills (2006); Mills and Healey (2006).

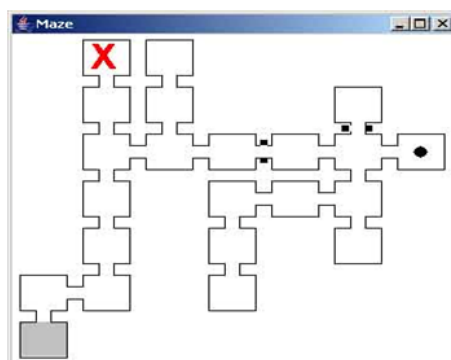


Figure 5.3: Solo Maze

5.2.3 Experimental Design

The experiments were run in 2 conditions: Normal and Downgrade. For each round of the experiment, the 3 participants were assigned to one of these conditions at random, resulting in a *between-subjects* design.

In the Normal condition there were no manipulations or interventions by the server. So all the participants saw the dialogue turns exactly as they were originally typed.

In the downgrade condition on the other hand, each dialogue turn sent by a member of the Duo team, is analysed in real-time by a module on the DiET server, checking if it contains any strong grounding cues such as ‘ok’, ‘yes’, ‘yeah’, ‘yep’ etc. In every other turn from the Duo detected thus, the cue is ‘downgraded’ to a provisional form of acceptance such as ‘erm’, ‘err’,

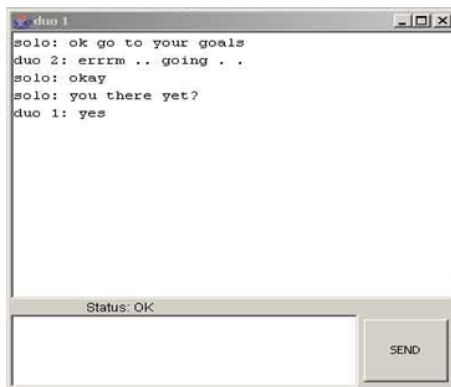


Figure 5.4: Duo 1 Chat Client

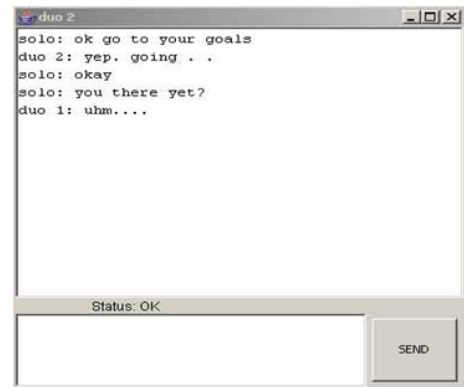


Figure 5.5: Duo 2 Chat Client

‘uurrm’ etc: the grounding cue is replaced by any one of these provisional acceptances at random, while the rest of the turn remains the same. The resulting turn is then sent to the other member of the Duo team instead of the original utterance. The Solo sees each turn as it is, unchanged; i.e. he isn't exposed to any of the downgrades (see Figures 5.4, 5.5 and 5.6).

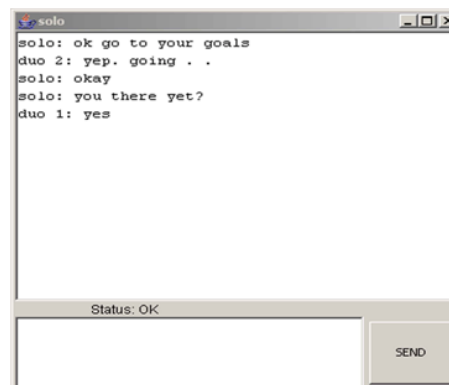


Figure 5.6: Solo Chat Client

The design of the experiment as described, is such that from the outset, the Duo are in a team, by virtue of the fact that they have identical mazes in all respects, and the same objectives locally and globally. At different stages, they need to describe their switch positions to the Solo (they have the same switches) and when they're at the receiving end, they **both** need to understand the location descriptions given by the Solo and act accordingly (they both need to go on his switch simultaneously to open the gates on the Solo's Maze). The duo thus have the opportunity to act for each other in their interactions with the Solo because they share some information (e.g. both know whether their gates are open or closed, both know where their switch points are). This also means that if one of them fails to do this, the others local objective (namely to open the gates for the Solo) is also compromised, and hence should (by hypothesis) try to repair the problem. The

downgrades of the strong grounding cues by the Duos create a situation where one Duo falsely believes that the other is having problems understanding or acting on what was just said by the Solo. As noted in the introduction, this creates a situation where a non-speaker of the moment (one of the Duos) is faced with the false belief that another (the other Duo), could not satisfy - via producing an appropriate response - the obligations imposed by the speaker's (the Solo) utterance.

5.2.4 Participants

84 participants were recruited from among undergraduate and postgraduate students from various disciplines at Queen Mary, University of London. 45 were male, 39 were female. The experiments took approximately 60 minutes to complete and each was paid 7 pounds for taking part.

5.2.5 Procedure

Participants were randomly assigned to the Duo and Solo roles. They were given written and oral instructions about the task together and had an opportunity to ask questions about the procedure. They were then seated in separate rooms, each at a desk with a PC displaying the appropriate task window (Figure 5.3 for the Solo and Figures 5.1, 5.2 for the Duos) and the chat client window (e.g. Figure 5.6). They solved as many mazes as they could in one hour and were then debriefed about the experimental manipulations together.

5.3 Hypotheses and Coding

The principal question for this analysis is whether the downgrades ('errm' etc.) of the strong grounding cues coming from one Duo (henceforth downgraded-duo) as seen by the other (henceforth non-downgraded-duo), have had any effect:

1. locally? Does the non-downgraded-duo do anything immediately to compensate for the apparent trouble the downgraded-duo is having in responding appropriately to what was just said? If so, what form does this compensation take? i.e. what are the devices used?
2. and globally? How do the downgrades impact the behaviour of the participants overall? This will be revealed through a comparison of the frequency of different phenomena in the normal and the downgrade conditions.

As noted, the duo have the opportunity to act for each other in their interactions with the Solo because they share some information (e.g. both know whether their gates are open or closed, both know where their switch points are). By hypothesis, if they are acting as a party, responses by one member of the Duo ought, all things being equal, be considered to stand proxy for responses by the other member of the Duo.

5.3.1 Initial Predictions

Before the experiment was carried out, the following predictions were made as to what the effects of the downgrades would be:

1. If one member of the Duo (apparently) downgrades their level of grounding this ought to degrade grounding for the Duo as a whole (note that only the other member of the Duo is exposed to the downgrade. The Solo receives the original, unaltered turn.). In other words we'd expect the downgrades ('erm', 'err' etc.) to be taken as collective by the non-downgraded-duo. The uncertainty created by the downgrade should, all things being equal, lead to fewer explicit acknowledgements by the Duo team in the downgrade condition, as the Duo would go on to repair the problem instead of producing an acknowledgement.
2. The Duos will, in order to deal with the uncertainty created by the downgrade, do more clarification interaction and ask more questions in the downgrade condition.
3. If we take a turn by a Duo to stand proxy for the other, then we expect that when grounding is apparently downgraded, the other Duo will compensate by producing a turn that he would not have otherwise. Therefore, we'd expect an overall increase in the number of turns by the Duos in the downgrade condition.

5.3.2 Post-hoc Analyses/predictions

In order to provide more support for the hypothesis and to rule out alternative explanations, after the experiment was concluded and the data analysis carried out with respect to the predicted patterns just mentioned, the following predictions were also made. The results of all the analyses (initial and post-hoc) will be reported together in the next section.

1. As discussed, grounding by proxy, if operative, entails that if one party member fails to ground a contribution, this would, in many instances, lead to direct interaction between or

‘conferring’ among the party members, in order to repair the problem. Thus, we’d expect the downgrades to lead to ‘extra’ direct interactions, or conferring among the Duo team in the downgrade condition, which means that we would expect more cases of explicit address by a Duo to the other Duo in the downgrade condition.

2. Recall case 2, in the introduction, of the ways in which, Lerner (1993) argues, parties may be projected: the use of the first-person plural pronoun ‘we’ where the scope of the pronoun is the members of the party. If the Duo are acting and see themselves as a party, we would expect more frequent uses of this pronoun by the Duo than the Solo overall, i.e. we’d expect an effect of task role on the frequency of use of this pronoun.

Importantly, in order to show that the Duo are acting as a party, we also need to show that the Duo behave differently overall, and that the effect of the downgrades on the Duos’ behaviour is different from that on the Solo’s; specifically that the Duo compensate differently and more than the Solo in downgrade condition, in terms of the above. Note that Solos weren’t directly exposed to the downgrade. We would nonetheless expect the downgrades to have an effect on their behaviour too, as the Solo will be exposed to the change in the behaviour of the Duo caused by the downgrade.

5.3.3 Coding

In line with the above predictions, each turn in the dialogue transcript was coded for the following dialogue phenomena:

- **Explicit Address:** Turns were coded for whether they contained explicit address cues (names of the participants). In particular if the turn contained an address cue, it was coded for whether the turn was addressed to the Solo or to a Duo.
- **Explicit Acknowledgements/Grounding Cues:** Turns were coded for whether they contained such cues. These were essentially the same cues that were downgraded.
- **Question Marks:** Turns were coded as to whether they contained any question marks (multiple question marks in the same turn count as 1).
- **Clarification Requests (CR):** Turns were also coded for whether they were explicitly asking a clarification question about a previous turn.

- The use of the first-person plural pronoun ‘we’: All turns were coded as to whether they contained this pronoun. In addition, the uses of the pronoun were coded for scope, i.e. who they were referring to. This was almost always quite clear from either the content of the turn being considered, or from the preceding or following turns, as the structure of the task and the arrangements of the Mazes were also known at the time of analysis. For example, the ‘we’ in “we need you to open our gates first”, cannot be referring to all 3 but refers to the Duo exclusively. In ambiguous cases, as noted, the turns that preceded the use of the pronoun and those that followed it, revealed what the participants had taken the scope of the pronoun to be. The uses of this pronoun had generally 2 scopes:
 - **Duo Scope:** ‘we’ used by a member of the Duo to refer to both of them, e.g. “we [the Duo] need you [the Solo] to open our gates now”
 - **Triadic Scope:** Used both by the Solo and the Duo to refer to all 3 of them, e.g. “we have to come up with a system”

5.4 Results

There were a total of 28 Triads: 14 in the downgrade condition and 14 in the normal. Two triads were removed from the analysis. Triad 4 was cut off mid-experiment as the participants had to leave. Triad 17 required additional instruction and clearly didn’t understand the task. They are therefore ignored. This left 26 triads, 13 in each condition.

Exact probabilities are reported but in all cases a criterion level of $p < 0.05$ was adopted.

5.4.1 Local Compensatory Devices: Some Examples

There were a few instances of the downgrades which had explicit, observable local effects. In particular, the downgrades had an explicit local effect if the downgraded acknowledgement was stand-alone, that is, when the whole turn was just an acknowledgement. This is expected, since turns which contain acknowledgements followed by something else, say a description, were actually ‘doing the work’ even when the acknowledgement contained within the turn was downgraded. For example: “Ok go to your goal now.” would become “errr go to your goal now”, which still doesn’t really manifest an acute problem. So the other Duo who sees the turn as downgraded would be unlikely to do much about it.

Below are some examples of the explicit, local compensatory devices employed by the non-

downgraded-duo, in ‘assisting’ the downgraded-duo in collectively carrying the conversation forward.

Clarification

- | | | | |
|-------|-----|--------------|------------------------------|
| (5.4) | (1) | Pak (Solo): | Im constrained to the top |
| | (2) | Dan (Duo): | Ok. [Bhups sees this as uuh] |
| | (3) | Bhups (Duo): | What? [to Dan] |

Paraphrase

- | | | | |
|-------|-----|-------------|---|
| (5.5) | (1) | Zee (Solo): | You gotta tell me where your other switch is. Open my gate so I can go find another switch. Because the other switch might be inside my gate. |
| | (2) | Sukh (Duo): | Ok. [Sim sees this as erm . . .] |
| | (3) | Sim (Duo): | Sukh, I think we have to tell Zee where our second switch is |

Repeat/Emphasis

- | | | | |
|-------|-----|--------------|----------------------------------|
| (5.6) | (1) | Bhups (Duo): | Dan, shall we go to goal |
| | (2) | Pak (Solo): | wheres dan |
| | (3) | Pak (Solo): | go to goal now |
| | (4) | Dan (Duo): | yes [Bhups sees this as errr] |
| | (5) | Bhups (Duo): | dan, Im going to the goal |

Elliptical turn assignment

- | | | | |
|-------|-----|------------|---|
| (5.7) | (1) | Milanista: | ilhaam are yuour gates row 5 C3, and R4 C3? |
| | (2) | Illy: | Ye [Malinista sees this as errr . . .] |
| | (3) | Milanista: | not gates sworry, switches |
| | (4) | Milanista: | grey boxes |
| | (5) | Andz: | erm... mine are |
| | (6) | Milanista: | ilhaam?? |
| | (7) | Illy: | I said ye if you read |

As is evident, all of the above are instances of pair-wise direct interactions between the Duos. As we will see below, this is in fact a systematic effect of the downgrades.

5.4.2 Descriptive Statistics

Before we present the statistics, some terms need defining:

A Maze Move is a task-level event, referring to the *successful* moving of the cursor - indicating the participant’s current position on the maze - by a participant, using the arrow keys on

	Normal Condition	Downgrade Condition
Total Number of Maze Moves	11686	9745
Total Number of Turns	5986	6179
Average Number of Mazes Attempted	9.31	7.38
Average Turns per Maze	49.45	64.4
Average Moves per Maze	96.55	101.57
Average Moves per Turn	1.95	1.58

Table 5.1: Descriptive Statistics

the keyboard, from one square on the maze to another *adjacent* one, e.g. moving the cursor two squares to the left and one up counts as three Maze Moves.

A (dialogue) Turn is a sequence of characters typed, followed by the return key, at which point the turn is packaged and sent to the server for processing (usually dispatched immediately to the rest of the participants).

Table 5.1 shows the overall descriptive statistics for the basic performance and task measures.

Not included in that table is, relevant to our 3rd initial prediction above, that there was a 3% increase in the total number of turns produced by the Duos in the downgrade condition, while there was a 3% drop in the number of turns by the Solos in the downgrade condition. But as we will see shortly, this difference is not statistically significant because of the high variability of number of turns across different triads and mazes. Nevertheless, it is a trend in the right direction, suggesting that the Duos work harder than the Solo to compensate for the uncertainty created as a result of the downgrades.

5.4.3 Basic Contrasts of Normal and Downgrade Conditions

Performance and Task Measures

For the parametric tests of task performance triad 57 were also excluded as an outlier on the basis that the total number of moves they made was more than two standard deviations from the mean thus significantly reducing the statistical power of the parametric tests. Note that they are included for the non-parametric tests on response types reported below.

Inspection of the frequency distribution of average moves per maze by each participant showed a positively skewed distribution (Skewness = 1.68). To reduce any distortion from this effect \log_{10} of moves per maze (Skewness = 0.47) were used for inferential analysis. This showed a main effect of Condition ($F_{(1,71)} = 9.41, p = 0.00$) but no main effect of Role ($F_{(1,71)} = 0.00, p = 0.96$) and no Role \times Condition interaction ($F_{(1,71)} = 0.00, p = 0.95$). Participants in the Downgrade condition made on average 17 moves per maze (SD = 14.7) whereas participants in

	None	Grounding Cues	Totals
Downgrade Condition	5349	830	6179
Normal Condition	5026	960	5986
Total	10375	1790	12165

Table 5.2: Explicit Acknowledgements/Grounding Cues

the Normal condition made on average 7.1 moves per maze (SD = 4.13).

The number of turns used by each individual on each maze was averaged and entered in a two factor analysis of variance with Condition (Downgrade vs. Normal) and Role (Duo vs. Solo) as between-subjects factors. This showed a main effect of condition ($F_{(1,71)} = 9.37, p = 0.00$) but no main effect of Role ($F_{(1,71)} = 0.78, p = 0.38$) and no interaction ($F_{(1,71)} = 0.00, p = 0.97$). In the Downgrade condition participants used on average 22.5 turns per maze (SD = 13.6) whereas in the Normal condition participants used on average 16.76 turns per maze (SD = 9.13).

Frequency of Explicit Grounding Cues in Normal Vs. Downgrade Conditions (Table 5.2):

The overall frequency of explicit grounding cues in turns in the Downgrade condition was slightly lower (13% of turns) than in the Normal condition (16%). This difference was reliable ($Chi^2_{(1)} = 16.44, p(\text{two-tailed}) = 0.00$).

Frequency of Use of Question Marks in Normal vs. Downgrade Conditions:

As noted, turns were coded for whether they contained one or more question marks (multiple instances were scored as 1). There is no reliable difference in the overall frequency of use of question marks in turns in the Normal and Downgrade conditions ($Chi^2_{(1)} = 0.69, p(\text{two-tailed}) = 0.79$, but see below for the frequency of use of question marks across different roles).

Frequency of Use of Clarification Requests (CR) in Normal Vs. Downgrade Conditions:

No reliable difference was found in the use of CRs in Normal vs. Downgrade conditions ($Chi^2_{(1)} = 4.10, p(\text{two-tailed}) = 0.52$).

	No Address	Address	Totals
Downgrade Condition	5420	759	6179
Normal Condition	5374	612	5986
Total	10794	1371	12165

Table 5.3: Frequency of Explicit Address

Frequency of Use of Explicit Address in Normal vs. Downgrade Conditions (Table 5.3):

As noted, inspection of the dialogues also led to a post-hoc hypothesis that there would be more frequent use of explicit address in the downgrade condition as participants work to re-establish common ground by making it explicit who their turns are addressed to. The frequency of use of explicit forms of address in the Downgrade condition was slightly higher (12% of turns) than in the Normal Condition (10% of turns). This difference was reliable ($Chi^2_{(1)} = 12.89$, $p(\text{two-tailed}) = 0.00$).

Overall, in the Downgrade condition there are fewer explicit grounding cues, no differences in use of question marks and slightly more frequent use of explicit address. These results show that the manipulation had an overall effect. However, in order to test the grounding by proxy hypothesis we need to look at the specific effects of the manipulation on the different roles.

5.4.4 Comparing the Effects of Roles

Several additional analyses were carried out to test whether the the downgrade manipulation had a different impact on the Duo and Solo roles.

Frequency of Use of Explicit Grounding Cues by Duo and Solo in Normal vs. Downgrade Conditions (Table 5.4):

In the normal condition, each of the three participants produces roughly the same number of explicit acknowledgements or grounding cues, e.g. ‘ok’, ‘yes’, ‘yeah’ etc. In the Downgrade condition this distribution changes. Solos continue to produce essentially the same number of grounding cues (321 vs. 324) but the members of the Duo produce significantly fewer (639 vs. 506, $Chi^2_{(1)} = 6.05$, $p(\text{two-tailed}) = 0.02$). There is thus a significant drop in the frequency of explicit acknowledgements / grounding signals by participants in the Duo role, in the downgrade

	Duo	Solo	Totals
Downgrade Condition	506	324	830
Normal Condition	639	321	960
Total	1145	645	1790

Table 5.4: Explicit Acknowledgements/Grounding Cues

	Duo	Solo	Totals
Downgrade Condition	345	324	669
Normal Condition	450	207	657
Total	795	531	1326

Table 5.5: Use of Question Marks

condition.

Frequency of Use of Question Marks by Duo and Solo in Normal vs. Downgrade Conditions (Table 5.5):

In the normal condition all three participants use roughly equal numbers of question marks. However, in the Downgrade condition the balance changes with Duo participants producing reliably fewer explicit questions and the Solo participants producing more ($Chi^2_{(1)} = 39.54$, $p(\text{two-tailed}) = 0.00$).

Frequency of CRs: There is no reliable difference in the frequency of CRs produced by Duo or Solo in the two conditions ($Chi^2_{(1)} = 0.06$, $p(\text{two-tailed}) = 0.81$).

Frequency of Use of Explicit Address Cues Taking into account who the turn is addressed to:

Post-hoc comparison of the frequency of use of explicit forms of address by the Duo and Solos in the two conditions reveals no significant difference ($Chi^2_{(1)} = 2.93$, $p(\text{two-tailed}) = 0.09$). This means that the overall increase in the use of explicit address in the downgrade condition is evenly distributed across the Duo and the Solo Roles.

However, we have not so far considered who the addressee of the turns containing the address cues are. Table 5.6 shows the distribution of address cues across conditions, according to whether they were produced by the Duo or the Solo and whom (what role) they were ad-

	Duo	Solo	Totals
Downgrade Condition	295	366	661
Normal Condition	203	310	513
Totals	498	676	1174

Table 5.6: Explicit Addressees By Role

dressed to. An additional post-hoc analysis shows that there are reliably more Duo to Duo address cues in the downgrade condition although this is a more marginal effect (Fisher's exact test, $p(\text{one-tailed}) = 0.04$).

Note that when the Solo uses an address cue she would always be addressing a (single) Duo. If we consider this together with the increase in the Duo to Duo address cues in the downgrade condition, we can infer that the downgrades result in more direct interaction between the three individual participants.

Frequency of Use of the Plural First-Person Pronoun 'we' by Duos vs. Solo:

To test whether Duo participants produce more instances of 'we' than Solo participants a Mann-Whitney within-subjects test of the frequency of use of 'we' by each participant was carried out. This showed a significant effect (Mann-Whitney $U = 151$, $p = 0.00$). Each Duo participant produced an average of 7.23 uses of 'we' (SD = 4.3) in contrast each Solo participant produced an average of only 2.1 uses of 'we' (SD = 1.3).

To test whether the Duo participants use more 'we's with scope over only the Duos vs. 'we's with triadic scope over a within-subjects Wilcoxon Signed Ranks test was performed on the number of times each participant's produced a 'we' of each kind. This showed a reliable difference ($Z = -5.19$, $p = 0.00$). Duo participants produce on average 5.7 (SD = 4.0) 'we's that refer to the Duo and 1.7 (SD = 1.7) that refer to the triad as a whole.

To sum up then, the post-hoc analysis of the use of the plural first-person pronoun, shows a large and reliable predominance in the use of 'we' by Duo participants compared to Solo participants. Moreover, there's also a large and reliable predominance of 'Duo Scope' over 'Triadic Scope' 'we's by the Duos.

5.5 Discussion

There seems to be clear evidence that the downgrades affect the behaviour of participants in both the Duo and the Solo roles.

Measures of task performance show that overall, the participants performed worse in the downgrade condition than in the normal condition. In the downgrade condition, participants in both the Duo and Solo roles, made significantly more maze moves on each maze attempted. They also produced significantly more dialogue turns during each maze. This suggests that the net effect of the downgrades was to make the participants work harder to complete each maze. There was, however, no Role \times Condition interaction, meaning that the participants in Duo and Solo roles, were equally affected by the downgrades.

On the measure of number of turns and maze moves, then, there is no difference in how the downgrades affect the participants' behaviour in Duo and Solo roles. This is not inconsistent with the grounding by proxy hypothesis, as the differences between the behaviour of participants in these roles, are reflected in the types of turn produced by Duos and the Solo, rather than merely the number of turns produced.

5.5.1 Failure to Ground by One Duo Depresses the Level of Grounding by the Duo Team as a Whole

Duo participants follow the predicted pattern. The drop in use of explicit grounding cues in the downgrade condition is exclusively due to changes in the behaviour of the members of the Duo. If one member of the Duo appears to ground less this has the effect of depressing the level of grounding by the Duo as a whole. The behaviour of the Solo is, on this measure, unaffected. This provides evidence that if one member of the Duo displays apparently weaker levels of grounding then the other member does too. The downgrades are taken by the Duos as collective.³

In effect, the uncertainty created about one participant's grounding level by the downgrades manipulation becomes shared by the other (Duo) participant. But, even though this is then transmitted to the Solo (although they aren't exposed to the downgrades, they are exposed to the overall reduction in grounding cues they cause) this does not affect the Solos use of explicit grounding cues.

As noted, this reduction in level of grounding does not, as we might intuitively expect, lead to an increase in explicit queries and checks of position by Duo participants (as indexed by Question marks). These turn types are, in fact, less frequent for Duo participants. In addition, there is no difference in their frequency of use of clarification requests, either in the downgrade condition

³Note that the opposite could be true - they could compensate by taking the initiative more, asking more questions, doing more explicit grounding.

overall nor by the Duo participants relative to the Solo participants in the downgrade condition. This suggests that where grounding appears to become uncertain for one of the Duo participants the other participant does not try to compensate for this through more explicit repair. This suggests either a reliance on more ‘indirect’ mechanisms for dealing with, or possibly avoiding these problems. One possible explanation for this unexpected effect is this: the ‘extra’ coordination between the Duo provoked by the downgrades means that potential questions and coordination problems which in the normal condition arise later on between the Duos, never arise explicitly in the downgrade condition, as they are addressed/answered as a result of the downgrades, earlier from the outset without having been raised/asked explicitly; hence fewer questions by the Duo in the downgrade condition.

5.5.2 Patterns of Direct Interaction

The tendency to use explicit forms of address is not different for the Duo or Solo participants. The overall increase in explicit address in the downgrade condition is therefore evenly distributed across the task roles. However, if we take into account the addressees of these turns, then we find, as predicted, that in the downgrade condition, the Duos address one another explicitly more frequently than in the normal condition. The Solo, too, addresses each individual Duo more frequently in the downgrade condition. This means, as noted, that the downgrades result in more direct pair-wise interactions between the three individuals.

This suggests that under ‘normal’ circumstances, that is, in the absence of grounding failures, a Duo can ground on behalf of the other. However, all things being equal, as grounding is degraded for one of them, this *temporarily* breaks the Duo team apart such that both the Solo and Duo participants treat each Duo as a single individual rather than the two of them as a party, leading to more pair-wise grounding between the Duos. As the uncertainty created by the downgrades becomes shared by the Duos and subsequently revealed to the Solo, leads to more pair-wise grounding between the Solo and each Duo separately. The Duos try to re-establish intra-party grounding, as predicted, via extra ‘conferring’ or direct interactions between themselves; and the Solo tries to re-establish grounding between himself and each Duo separately and individually. This supports the grounding by proxy hypothesis, in that in the absence of negative feedback from ‘the proxy’, grounding seems to operate between the party as a whole and those outside with whom the party is interacting. Failures to ground by a member however, lead to more pair-wise grounding between the party members themselves and between each individual

member and those with whom the party is interacting. The party is ‘broken apart’ *temporarily*, as a result of the downgrades, both from the perspective of the party members themselves (the Duos) and those outside (the Solo).

Note that all the local examples except the first one, presented above, also contain Duo to Duo address cues.

Considered together, there is evidence that one important configuration of the participation framework, in which the mutual grounding responsibilities between the members of a party are made manifest, is, as predicted, through such ‘extra’ *conferring* or direct *intra-party* interactions or grounding among the members.

5.5.3 The Duo Act and See Themselves as a Party

As noted in the introduction, one of the ways in which parties are projected is via the use of the plural first-person pronoun, with scope exclusively over the party members. We hypothesised that this should also be one of the ways in which parties already projected and relevant are made manifest.

The significant difference between the frequency of use of the plural first-person ‘we’ by the Duo and Solo, together with the predominance of Duo scope ‘we’s by the Duos in both conditions, clearly shows that the Duo are seeing themselves and acting as a single party. On many occasions, an utterance by a Duo, in which this pronoun (or the possessive/objective counterparts, namely, ‘our’/‘us’) is used, is not repeated by the other Duo (*proxy in speaking* as per the introduction). The relevant information is already conveyed through one utterance alone. That is, these utterances stand proxy for whatever the other Duo has to say (e.g. “our switch is on the top left”). Also, where such utterances are produced in response to an utterance by the Solo, this is also taken to stand proxy for the other Duo’s response (e.g. “Solo: Did your gates open? Duo 1: no. we’re still stuck.”, *proxy in response*). Often, the other Duo would not say anything as it would clearly be redundant. Note also, that since the Duo are publicly known by all to have always exactly the same maze and the same maze/gate status, the use of any form of this pronoun is often not even necessary; for example: “Solo: Did your gates open? Duo 1: no still closed”. The other Duo in this case would often not repeat this, as the utterance equally applies to her maze.

This result supports the hypothesis that the Duos act as a unified aggregate, or a party, such that an utterance from one member, whether in response to a Solo’s first pair-part, or as a first

pair-part itself, can stand proxy for what the other Duo has to say.

5.5.4 Solo's Compensation: Irreducible Collective States

The Solo participants' grounding behaviour, in terms of explicit grounding cues, doesn't appear to change in the downgrade condition. However, it is clear that they are sensitive to the manipulation since their use of question marks reliably increases. This is a situation in which the effect of the downgrades is propagated, such that the Solo is affected indirectly, partly as he loses 'confidence' in the Duo as a team, because of their 'extra' coordination problems, lack of explicit grounding, indecision and conferring (which, unlike the downgrades themselves, are visible to the Solo). Moreover, the downgrades affect the way the Solo treats the Duo. Whereas in the normal condition he seems to treat them as a party, interacting with them as a whole, in the downgrade condition, he treats them more as individuals, thus addressing them individually.

If such dialogues resulted in states of understanding, reducible to their component dyadic interactions, we would expect the Solo not to be affected at all as he is not directly exposed to the downgrades. This shows that such multi-party dialogues result in dialogue contexts that are more than can be predicted by considering the prior dyadic interactions alone, thus providing experimental support for our conclusion about the resulting irreducible collective contexts or states of understanding in the previous chapter.

Considered together, these findings strongly support the grounding by proxy hypothesis, namely that the Duo act as a party and that the 'principal axis' of the dialogue is between the Solo and the Duo as a whole rather than between the three individual participants; i.e. the results indicate that grounding, like turn-taking, could, modulo grounding 'failures' by 'the proxy', operate not between individuals but between parties as higher order entities. These results also constitute causal evidence, hitherto absent from the literature on multi-party dialogues, for the practical reality of parties.

5.6 Conclusion

This chapter constitutes an attempt to operationalise the Principle of Responsibility coupled with Schegloff's notion of parties in multi-party conversations. We proposed that in conversations where 2 or more participants form a party with respect to the local communication task at hand, the party as a whole is responsible to carry the conversation forward, and that its members act as a unified aggregate in doing this. In this manner, one member can *stand proxy* for the other

members in grounding/carrying the conversation forward, while the rest of the (momentarily silent) members are at no disadvantage in terms of grounding: Grounding by Proxy. Accordingly, the contextual increments arising from a dialogue move/action by one party member, can, in the absence of negative feedback from other members, affect the party as a whole.

We provided causal, experimental evidence for the practical reality of parties, characterised via the notion of Grounding by Proxy. We showed experimentally that this entails a set of mutual grounding responsibilities between the members of a party. Namely, that on the occasion that one of the members fails to satisfy the constraints imposed locally by the current context of the conversation, then the others will compensate in doing this as this failure to ground becomes shared by the party as a whole. The conversation cannot move forward until the issue is resolved among the party members. This is done either through more direct pair-wise grounding among the party members or through direct pair-wise interactions between those with whom the party is interacting and each individual party member. We highlighted a set of devices that the party members might use to re-establish intra-party grounding and the corresponding configurations of the participation framework in which this may occur.

Furthermore, our conclusion from the previous chapter, namely that such dialogues lead to collective states of understanding not reducible to their component dyadic interactions, was supported experimentally.

Importantly, the results presented here suggest that if conversational responsibilities are to be indexed to participatory status in multi-party dialogues, we need to distinguish between the status of the (ratified) participants who fall inside and those who fall outside of the parties locally projected, active and interacting, and that this should determine the level of understanding/acceptance that these participants, in withholding of feedback, are assumed to have reached. Let's call these primary and peripheral participants respectively. We will have occasion to refer to them as such in the rest of this thesis.

But note that in normal, everyday conversation, the parties determined thus, are in a constant state of flux as the conversation moves forward, since party formation seems to be strongly tied to topic and every topic is potentially different. In the grounding criterion: "... to a degree sufficient for current purposes" (Clark & Schaefer, 1989), the 'degree' is therefore in a constant state of flux, in terms (at least) of who needs to give evidence of understanding/respond. Whether party formations are really tied to topic as suggested, is one of the themes of subsequent chapters. We will not dwell any more on this here.

Note that what we've argued in this chapter does not really answer the question of what is constitutive of parties. What are the necessary relations that should hold between the information states of a set of participants, and a given topic for them to be regarded as belonging to a single party with respect to that topic? This is an important open question that has not been addressed here. We have not, in other words, defined parties, but have assumed rather that they are projected and relevant, and have analysed the effects or manifestations rather than the reasons or causes of party formations.

Chapter 6

Distinguishing Dialogue Contexts

6.1 Introduction

Theories of human interaction use the notion of ‘conversation’ as a basic building block. It is supposed to group together the sets of people, utterances, gestures or other communicative signals which are considered together in an analysis. In the case of two-person, dyadic interactions this choice is relatively unproblematic. However, in gatherings of three or more people the boundaries between conversations become both more important for understanding the structure of the unfolding interaction and more difficult to track.

In section 3.2.4, we elaborated on how the methodological difficulty of characterising the boundaries of conversation as a unit, has lead researchers to focus on what Goffman (1963) has called ‘focused interactions’. Conversation as a unit has been equated with focused interaction (e.g. (Goodwin, 1981)), and delineated typically in terms of physical criteria that are intended to capture who the ratified participants are. In this regard, we highlighted Goffman (1981)’s definition of conversation in terms of physical orientation, movement and the use of ‘ritual brackets’, which can mark the boundaries of interaction. Furthermore, we alluded in several places to Kendon (1990)’s descriptions of more subtle ways in which people use body position and orientation to maintain interactional units or ‘F-formations’. An F-formation is also, effectively, a ratificational device (recall the discussion of ratification in section 3.2.8).

Nevertheless, as exemplified in section 3.2.4, the characterisations of conversation as a unit in terms of the set of ratified participants, alluded to above, are contradicted in discontinuous strands of interaction when conversational contexts are ‘left open’ - questions left unresolved -

and taken up later on; in the sense that what you say later on, counts as ‘pragmatically next’ or a second pair part, to what was said before the conversation was ‘broken’, e.g. in multi-focus gatherings where one person may be a participant in more than one parallel conversation. This will become clearer shortly below when we look at some more corpus examples.

In this chapter, we present experimental evidence that even within a single, focused multi-party conversation, more than one dialogue context can emerge as a result of fluctuations in different participants’ levels of involvement. These contexts, we propose, are more fine-grained than any ‘interactional unit’ defined solely in terms of who the ratified participants are and are also independent of topic.

As we will discuss in more detail in the final chapter of this thesis, our findings here, have important ramifications for formal and computational theories of discourse and dialogue that aim to characterise syntactic and semantic dependencies across multiple turns and resolve the meaning of elliptical expressions accordingly (e.g. Ginzburg’s KoS sketched in section 2.7.1). In order to achieve this, it is critical for these theories to be able to specify what constitutes the shared context for a conversation.

6.1.1 Ellipsis as window on shared context

As we showed in chapter 4, ellipsis provides an interesting and more fine-grained window - than participants’ physical orientation to one another as for example expanded by (Kendon, 1990) - on when/whether context is shared for the parties to an interaction. In dialogue this context is often another turn/utterance: the *Antecedent*. Interpretability of elliptical utterances - as manifest in the relevant response to them - demonstrates therefore that the context required to recover the missing bits (the antecedent) is sufficiently *Salient* or one might say pragmatically ‘close’ or ‘in focus’, for the responder.

In dyadic dialogue, the antecedent of an ellipsis is almost always found in the immediately preceding turn (Purver, 2004) e.g. A: Where did you go yesterday evening? B: Cinema. In multi-party dialogues however, the situation is more complex. Consider dialogue 1 from the AMI corpus below:

- (6.1) (1) C: What does cutest spelling mean?
 (2) B: oh, she spelled cutest um with an L,
 (3) C: oh, okay
 (4) B: so that that's just something I pointed out
 (5) D: oh yeah
 (6) A: **Cutest?** [Clarification Ellipsis, Antecedent Distance of 5]
 (7) D: E.S.T_
 (8) A: Thank you

Dialogue 1: From the AMI corpus (Carletta, 2006)

In the above example, A's elliptical request for clarification (turn 6) is, in effect, just as local to C's utterance as B's (at turn 2). As highlighted in section 3.10, the adjacency structure of multi-party dialogues is thus generally more complex than dyadic dialogues. In dyadic dialogues, exchanges (adjacency pairs) are comprised of one initiating and one reacting move (e.g. question/answer or assertion/acceptance). In multi-party dialogues on the other hand, after the initiation of any move, multiple non-speakers could react to it elliptically without necessarily causing any trouble in understanding. What this means is that adjacency structure in multi-party dialogues is not linear, but instead looks more like a tree: Given any utterance, there can be **more than one** subsequent turn, constructed as 'pragmatically next' to that utterance.

We can also find examples in which multiple conversational strands, with overlapping groups of participants, are interleaved. Consider Dialogue 2 below which comes from a corpus of on-line chat community interactions in an environment called Walford (described by (Healey, White, Eshghi, Reeves, & Light, 2007)). The residents of Walford interact using a single text chat screen on which all the messages they send and receive are displayed. In Walford, people can conduct multiple concurrent conversations with different, potentially overlapping, groups of individuals and they routinely take advantage of this fact. These conversations all take place in the same chat window with no visual cues to help distinguish the different conversational strands or groups of participants. Dialogue 2 shows a sequence of turns seen by 'Naomi'. Note that we have used indentation to indicate the separate conversational tracks but in the actual chat window there is no indentation:

- (6.2)
- (1) Sonia: Hopws you?
 - (2) Naomi: tired
 - (3) Naomi: you?
 - (4) Sonia: Just got in from work
 - (5) Naomi: ahh I see, kinda sick and a bit pissed [Antecedent of CE]
 - (6) Naomi: heya hun
 - (7) Naomi: Logged in from home?
 - (8) Frank: ok, Im off
 - (9) Garry: yep, just about to start my studying
 - (10) Frank: Yeah Yeah Yeah
 - (11) Sonia: **Drunk pissed?** [Clarification Ellipsis(CE), distance of 6]
 - (12) Naomi: Angry pissed.

Dialogue 2: From the Walford Corpus

Naomi is conducting a dyadic exchange with Sonia, a second dyadic exchange with Garry - which Sonia does not see - in parallel with a third, broadcast ‘group’ interaction with Frank that is seen by everyone. Despite this mixture of turns and participants it is clear that the clarification ellipsis at turn 11 has its antecedent in turn 5. It is as if the conversation/shared-context between Naomi and Sonia is ‘suspended’ for the duration of turns 6-10, the intervening sub-dialogue. Sonia’s subsequent CE is then successfully understood by Naomi as referring back to turn 11.

6.1.2 Accessibility of the Antecedent

What these examples make clear is that the accessibility of an antecedent is not a simple function of time elapsed, turns elapsed, nor recency in memory. They suggest rather that people track different conversational contexts that may be maintained in parallel over surprisingly long periods especially in discontinuous strands of interaction where there is an ‘open state of talk’ (Goffman, 1981). An initial generalisation suggested by these examples is that distinct dialogue contexts may be distinguished in terms of the set of primary participants, i.e. those actively making contributions (jointly or individually). This leads to the prediction that an elliptical turn that is separated from its antecedent by a different dialogue context, i.e. when there’s been a change in primary participants, should be more easily resolved than one which is separated by the same number of turns from the same dialogue context.

In this chapter, we report an experimental test of this prediction by looking at the relative ease with which participants are able to resolve elliptical clarification questions within or across dialogue contexts, distinguished here, in terms of the set of primary participants.

6.2 Method

6.2.1 Task and Materials

The experiment uses a variation on the ‘Tangram’ (Chinese abstract figures) tasks used in various experimental studies of the common ground (e.g. (Wilkes-Gibbs & Clark, 1992)). In these studies, a *Director* has a sequence of pictures of these figures in a fixed order in front of him. The director describes these one by one to a *Matcher*. The matcher has the same figures, but unordered, which he needs to put into the same order as the director. They cannot see one another’s cards, but they can talk as much as they like about them.

3-Way Tangram Task

We use an on-line, 3-person variant of the same task. The task involves 2 directors and 1 matcher (D1, D2 and M). The directors share a set of 12 Tangram images in a single target order but half are assigned to one director and half to the other. At the start of the task all the figures are hidden from the directors but they can see which items in the sequence they will deal with and which ones the other director will deal with (see figure 6.1). The matcher, by contrast, can see all the figures but, as with the original task, does not know the target order they will have to put them in (see figure 6.2).

The directors proceed through the sequence of figures one at a time by clicking, as appropriate, in the window. For each figure the ‘active’ director provides a description so that the matcher can identify it and drag and drop it into the appropriate slot (see figure 6.2). The ‘passive’ director does not see the image and therefore has no direct involvement in describing it. They are, however, a fully ratified side-participant in the exchange. In this manner, for each round of Tangrams, we gather a sequence of interactions between D1 and M, and some between D2 and M with the sequence of participants controlled experimentally.

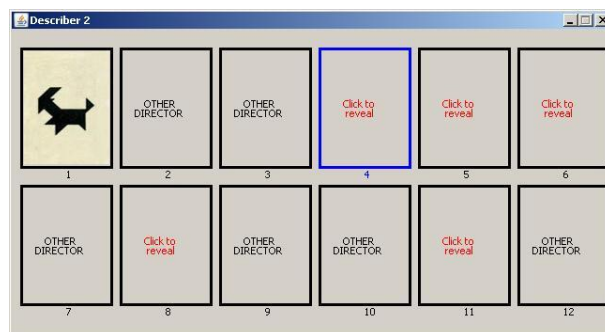


Figure 6.1: Directors’ view



Figure 6.2: Matcher's view

In order to enable the insertion of the experimental probe clarification questions all interaction between the directors and matcher is via the DiET Chat Tool, described in section 5.2.2.

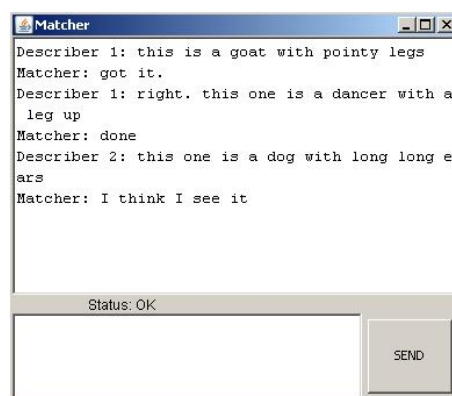


Figure 6.3: Chat Display

In this experiment, we parse utterances on the fly, in order to identify possible fragments from the directors turns that can serve as target sources/antecedents for for long-distance Reprise Fragments (RF, a special class of Clarification Ellipsis. E.g. A: Did Mary show up last night? B: Mary?[RF]). The criteria for selecting a potential target in this experiment were that it should be a noun phrase, verb phrase or prepositional phrase, that is not subsequently repeated prior to its use in a spoof RF (in order to factor out the element of ambiguity with respect to the current figure). Given a suitable target, the server can then generate a spoof clarification question that appears to originate from the Matcher, but is not relayed to the Matcher's screen. Both directors see the RF and either may respond. The response is trapped by the server and the dialogue resumes (see Dialogue 3 for an example).

- (6.3) (1) D1: the big fat lady. [Antecedent of RF]
 (2) M: ok got it.
 (3) D2: ok my one now.
 (4) D2: camel with the triangular back.
 (5) M: ye got it
 (6) M: sorry err fat lady? [Fake server probe (RF).]
 (7) D1: yeh that was number 2 [Trapped Response]

Dialogue 3: Example from the Experimental Corpus

This whole process can be carried out without disrupting the dialogue and without being detected by any of the participants. For a more detailed description of this technique see (Healey et al., 2003).

The probe RF interventions are triggered when the Matcher places a figure in the appropriate slot, and just before everybody moves on to discuss the next figure. The target antecedent of the RF is always from the last-but-one item, i.e. the fragment queries something said previously by one of the directors about the penultimate figure described so far.

6.2.2 Experimental Design

As noted, in the experiment reported here the probe spoof RF is always generated on an antecedent (target) expression that has been used to describe the *previous* Tangram figure in the sequence. The experimental contrast is between a same dialogue context condition where the current Director also described the last-but-one item (i.e. there has been no change in the primary participants over the two items) and a different dialogue context condition where the other Director described the last-but-one item (i.e. there has been a switch in primary participants over the two items). This created a situation in which probe RF's occurred, on average, at equivalent distances from their targets in terms of time and turns elapsed but at different distances in terms of dialogue context boundaries (within dialogue context in the former condition and across sub-dialogues in the latter). Exactly 4 RF's (2 in each condition) were sent to the directors per round of 12 Tangrams resulting in a within-subjects manipulation.

6.2.3 Participants

60 participants were recruited from among undergraduate and postgraduate students from various disciplines at Queen Mary, University of London. 33 were male, 27 were female. The experiments took approximately 60 minutes to complete and each was paid 7 pounds for taking

part.

6.2.4 Procedure

Participants were randomly assigned to the Director and Matcher roles. They were given written and verbal instructions about the task together and had an opportunity to ask questions about the procedure. They were then seated in separate rooms, each at a desk with a PC displaying the appropriate task window (Figure 6.1 for the Directors and Figure 6.2 for the Matcher) and the chat client window (Figure 6.3). They did as many rounds of tangrams that they could in 1 hour and were then debriefed about the experimental manipulations together.

6.3 Results

The principal question for our analysis is whether the dialogue context manipulation affects the responses Directors give to the spoof RF probe questions from the matcher. In particular, our prediction is that the distant antecedents should be *more accessible/salient* - and hence the RF, *easier* to interpret - when they are from a different dialogue context - because they are pragmatically 'nearer' for the Director who produced them - than when they are from the same dialogue context i.e. were produced by the director who is describing the current item. In contrast to this, if accessibility is solely a matter of recency or total amount of intervening turns then there should be no difference between these conditions.

Because both Directors saw each RF probe question both had the opportunity to respond regardless of who actually produced the target antecedent. Responses were therefore classified according to who responded: Owner (of the target expression), Not Owner (i.e. the other director) and None where nobody responded. As Figure 6.4 shows, the distribution of responses was different in the two dialogue context conditions. The probe question was equally likely to receive no response in the two dialogue context conditions but where the target crosses the dialogue context boundaries ('different dialogue context' condition) the owner was less likely to respond than the current speaker (this is construed as *Momentum of Dyads* in the discussion below). A multinomial regression analysis confirmed the reliability of this pattern $Chi^2_{(2)} = 46.41, p = 0.00$.

Where a response was given, it was also classified as to whether it displayed a *Local* (i.e. incorrect) or *Distant* (correct) interpretation, or no interpretation at all (the response itself was a request for clarification). This was possible, because the figures being described and the order in which they were described by the directors, were known at the time of analysis. In addition, the

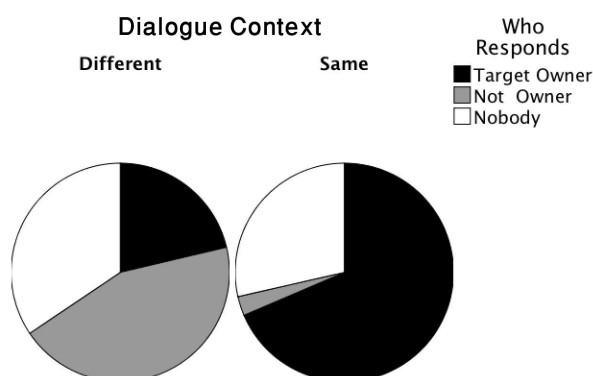


Figure 6.4: Distribution of responses across conditions

whole dialogue between Directors and Matcher prior to the sending of the RFs, was recorded. So the response often revealed clearly, whether the Director had taken the RF as clarifying something about the current figure (Local/incorrect), or the previous one (Distant/correct). In particular, the responses were coded as Distant, if:

- part of the original distant antecedent from which the RF was taken, was repeated by the Director.
- there was reference in the response to the previous item e.g. “D1: yea that was number 2”.

Otherwise, the responses were coded as displaying a Local interpretation.

The responses were equally likely to receive no interpretation at all in the 2 conditions. But, as Figure 6.5 and Table 6.1 show, where the responses did reveal either a distant or a local interpretation, the pattern of interpretations was systematically different in the two dialogue context conditions. If the owner responds then a distant interpretation is only favoured where the antecedent was in a different dialogue context. If the owner responds and the antecedent comes from the same dialogue context, i.e. was produced by the person who also produced the most recent description they strongly favour a local (incorrect) interpretation. This pattern is confirmed by a multinomial regression analysis which shows a main effect of both dialogue context condition $Chi^2_{(1)} = 18.26, p = 0.00$ and of who responds $Chi^2_{(1)} = 21.25, p = 0.00$ on the likelihood of a distant (correct) interpretation.

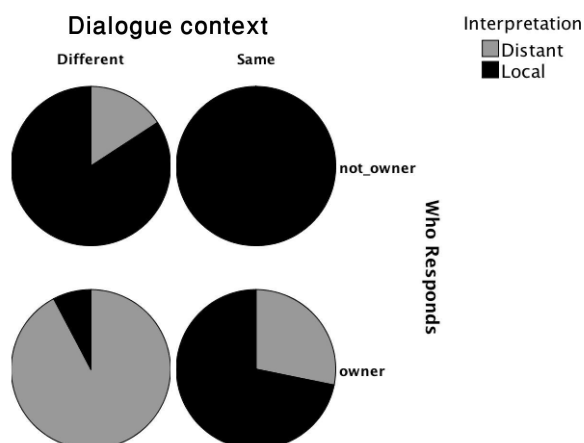


Figure 6.5: Distribution of the Directors' Interpretation of Probe Reprise Fragments Across Conditions

	Dialogue Context			
	Same		Different	
	Local	Distant	Local	Distant
Owners' Interpretations	28	11	1	12
Non-Owners's Interpretations	1	0	16	3

Table 6.1: Distribution of the Directors' Interpretation of Probe Reprise Fragments Across Conditions

6.4 Discussion

6.4.1 Strong tendency to interpret a fragment against local context

The most obvious pattern in the results of this experiment is that there is a strong preference for local antecedents for ellipses. Even where the owner of the distant antecedent responds to the probe RF they frequently interpret it as being about the current item and not the (distant) target. This is despite the fact that the fragment used in the probe question has not been repeated since. This suggests that, all things being equal, there is a strong presumption that ellipses will have local antecedents.

6.4.2 Momentum of dyads

In the 'same dialogue context' condition the current speaker coincides with the owner of the distant antecedent (the 2 consecutive figures were described by the same director). In the 'different dialogue context' condition, the current speaker would be different from the owner of the distant antecedent. As figure 4 indicates, there's a strong preference for the non owner (coincident with the current speaker) to respond in the 'different dialogue context' condition. In both con-

ditions then, the current speaker is more likely to respond. Therefore, a possible generalisation here is that all things being equal, by default, in the absence of any address cues, utterances are taken to be addressed at the last speaker; dyads therefore have momentum. This is consistent with the results found in (Parker, 1988). The significant finding in that study is that a random utterance in a small focused group discussion is most likely (by a large margin) to be within a dyadic turn-taking pattern (A-B-A). Even when this pattern is broken, a new one is quickly established and sustained for an extended period. This means that even interactions with a small set of fully ratified participants, tend to lapse into stretches of dyadic talk during which the rest of the participants take on secondary roles. We argue below, that it is exactly such imbalance in the participants' level of participation, which results in the emergence of more than one distinct dialogue context; hence the need to index shared-context to specific sets of participants.

6.4.3 Emergence of distinct dialogue contexts

Importantly for our purposes, the results show that dialogue context structure, understood here in terms of changes in primary participants, has a significant effect on the accessibility of the antecedent of an ellipsis, prior to the change of primary participants. As predicted from the corpus examples, the long distance antecedents are more easily accessed where they are from a different dialogue context. It's crucial here to see that it is not just any change in the primary participants across the sub-dialogues, that makes the distant antecedent more accessible. For the change in the primary participants to have such an effect, it is essential that:

1. The participant producing the elliptical turn, say *A*, and the owner/producer of the antecedent, say *B*, are *primary participants* in the antecedent containing, distant dialogue context i.e. they have grounded the antecedent.
2. Either *A* or *B* is NOT a primary participant in the intervening dialogue context.

Support for 1 above, comes from the result that the preference for the incorrect local interpretation was strongest when the non owner responded. This seems to be because the non owner is never a primary participant in the distant dialogue context containing the antecedent, and therefore has not grounded the antecedent to a level that would allow him to interpret the ellipsis as distant. So we may say here that the context that *A* shares with *B* prior to the intervening dialogue context (they are both primary participants in it), is 'suspended' as a result of either *A*'s or *B*'s lack of active participation in the intervening dialogue context, so that subsequent to it, that

context/antecedent is still sufficiently salient. And this is despite the fact that *A* or *B* may be a ratified side-participant to the intervening exchange.

Why is this happening one might ask? Despite the RFs containing often unambiguous cues, pointing to the distant context, the primary participants still show a tendency to interpret the RFs as local. There's therefore a contention between the contexts (local and distant) against which the primary participants assess these fragments. This shows that for the primary participants the local context or topic has priority, as providing the antecedents against which elliptical utterances addressed at them are resolved.

In contrast, this contention seems to disappear almost completely for peripheral participants. We suggest that this is because all the ratified participants mutually know that the peripheral participant of the moment, does not have any concern or responsibility with respect to what is being locally discussed and as noted that he doesn't say anything about the current referent. This means that in contrast to the locally primary participants who are strongly attached to the local context which they're collaboratively building up, the peripheral participants are only weakly attached to it. We might therefore say that the peripheral participants' level access to the distant context is high, because the local context is unavailable to them. As noted, their view of the dialogue context has not 'moved forward', while the primary participants' has.

All this supports the idea that multi-party dialogues can lead to conversational contexts that have a tree like structure. As a result utterances that are placed on either side of a sub-dialogue (corresponding to a single dialogue context in our sense) can be understood as pragmatically 'next' to each other. The experimental approach has allowed us to factor out from this effect the possible influence of number of turns, topics (in terms of Tangrams described) and the amount of time elapsed between the ellipsis and its antecedent. This provides a very useful notion of context locality/saliency as indexed to specific participants - contingent upon their level of participation - which accounts for the adjacency structure of turns, not only in multi-focus, discontinuous situations (such as Dialogue 2 above) but also focused gatherings with sets of fully ratified participants.

6.5 Conclusion

The experimental results presented here show that even within small groups of fully ratified conversational participants distinct dialogue contexts can emerge as a result of changes in the primary participants, and hence they are, as indexed by ellipses, anaphora and other context sen-

sitive expressions, more fine-grained than ‘interactional unit’s, such as an f-formation (Kendon, 1990), defined solely in terms of who the ratified participants are. Moreover, we have argued for the need to index these contexts to specific sets of participants and provided conditions, in terms of the participants’ statuses (peripheral vs. primary) during and with respect to a stretch of talk/sub-dialogue (corresponding to a single dialogue context), under which utterances on either side of that sub-dialogue, may be regarded as ‘pragmatically next’ to each other.

As we will see in more detail in later chapters, the results presented here imply that one and the same utterance can give rise to divergent contextual updates on the information states of primary vs. peripheral participants, and hence that the respective contextual update rules responsible for the utterance by utterance transitions of these participants’ view of the context of the dialogue, need to be distinct.

In any attempt to scale-up computational or formal models of dyadic dialogue, to multi-party dialogue, the patterns observed here need to be taken into account, so as to track and update as the conversation moves forward, what is salient for whom at any given point, and to evaluate the meaning of elliptical expressions accordingly.

In the final chapter of this thesis, we will sketch how the empirical results presented here may be construed within Ginzburg’s model of dialogue context, and present some tentative ideas as to how the model could be transformed so as to capture these results.

Chapter 7

Collective Vs. Dyadic Contexts

7.1 Introduction

Recall that in chapter 4, we showed that despite the locally unbalanced levels of feedback between the ratified participants (addressee vs. side-participant), common ground can accumulate evenly, in the same way, for all of them. Chapter 5 was an attempt to characterise the systematics of grounding in such conversations; that is, at least, those conversations in which some of the participants are organised into parties, in virtue of the relations holding between the participants' publicly known collective responsibilities, concerns and goals, and the local conversational topic. We proposed that ratified participants which, at various points, fall inside and outside such parties need to be distinguished, in terms of grounding responsibilities and correspondingly the level of understanding/acceptance that, in withholding of feedback, they are assumed to have reached.

In the previous chapter on the other hand, we presented experimental evidence that in a single multi-party conversation, under the same local conversational context, in terms of dyadic context boundaries, there's a systematic conflict between participants' interpretations of similar Reprise Fragments. This was owing to the distinctness of the contexts against which they were being assessed. This was the sense which the more fine-grained *dyadic* shared-contexts may or may not 'move forward' for specific pairs of participants, expressed there, in terms of the saliency of distant contexts/antecedents.

In this chapter, we will be addressing the question of how/whether in similar terms, a participant who falls outside a party for a stretch of talk, is distinguished from one who falls inside. In other words we'll be addressing the question, posed in identical terms as the previous chapter,

of the effects of collective vs. dyadic participation, on the saliency of the contexts immediately prior to when the new party was established and the old one suspended. Let's make this more specific.

In the previous chapter, we showed that the saliency of a prior context/antecedent/question among participants *A* and *B* is, independently of distance in terms of turns/time or topic, contingent upon whether they have, together, participated in the stretch of talk, since that antecedent was produced. This is the sense in which the $\{A,B\}$ shared-context does or doesn't move forward, as a result of the pair's *joint* participation and presumed responsibility, or lack thereof, since the distant antecedent/context.

Nevertheless, in the previous chapter the dialogue contexts were always dyadic. Therefore, questions as regards the effect of intervening collective contexts on the accessibility of distant antecedents prior to it, remain unaddressed, i.e. what if our previously peripheral, silent side-participants in the intervening contexts, were also to be an active, responsible, primary participant? What if, in other words, 2 or more participants were to act as a party in the intervening context? Wouldn't we be inclined, given the results from chapters 5 and 6, to say that the contexts that each member of a party shares with the participants outside - with whom the party as a whole is interacting - move forward together, evenly? If so, in light of the results from the previous chapter, this would entail that collective participation, in the stretch of talk that intervenes between a distant context/antecedent, prior to when a new party was established and the old suspended, should make that antecedent equally less salient/accessible for each member of the newly established party, as compared to those who are outside the newly established party, but who nevertheless grounded the antecedent (i.e. were a primary participant (a member of the old party) in the antecedent containing context). This is, partly, what we aim to show experimentally in this chapter.

Recall that in the conclusions to chapter 5, we pointed out that it is as yet unclear whether party formations are really tied to topic. This is another issue that will be addressed here. In particular, we will show experimentally that upon topic shifts new parties may be established and the old ones suspended or destroyed, i.e. that the participants' arrangements in the active, interacting parties could change upon topic shifts (see below for a further discussion of what is meant here by topic and how it relates to the participants and context).

As will become apparent shortly, the dialogue contexts that we claimed can emerge in the last chapter, can also be understood to emerge as a result of one participant falling outside a party

- and hence becoming an inactive/indifferent peripheral participant - for a stretch of talk. Our experimental results indicate, in line with those of the previous chapter, that this locally peripheral participant is very weakly attached to the local context, which is being collaboratively built up by the locally active party members. This results in the higher accessibility of the prior distant context - before the new party was projected/formed - for this locally peripheral participant - who was a member of the old party which has been suspended or destroyed in the local context - as compared with the locally primary participants. Thus, as noted above, another way in which the participants outside and inside the interacting parties are distinguished is in terms of the contexts/antecedents that others' elliptical utterances received by them get assessed/resolved against. This, we will argue, is evidence that those who fall outside all the interacting parties and those inside, that is, peripheral and primary participants respectively, effect systematically divergent contextual increments/updates, as a result of one and the same utterance. This gives rise to the systematic emergence of distinct dialogue contexts, which, all things being equal, were caused by changes in the arrangements of the participants in the active, interacting parties. As noted, this is, in turn, caused by topic shifts, whereby the new topic projects or makes relevant different parties than the old ones.

7.1.1 'Topic' as indexed to the participants?

In the experiment to be described, analogously to that in the previous chapter, the participants will be required to make a series of references (Tangrams). The shift from describing one referent to the other is essentially what we've described as 'topic shift'. In the previous chapter, we claimed that the dialogue contexts that we showed emerge are *independent of topic* (Tangrams). With the notion of topic being under-explicated in general, and in particular that of topic shift as yet to a great deal underspecified and problematic in the literature (nobody really knows what constitutes a topic shift), such parallel between a shift in the referent being described and a shift in topic may seem contentious.

Nevertheless, the claims that have been made are not about the effect of topic shift per se, but that of the level of involvement of those present and hence their participant statuses *as determined by* the relations holding between them and what is being talked about (topic). In other words, what we have been and are concerned with here is the effect of shifting parties, group structures and presumed conversational responsibilities, on the evolution of dialogue contexts. The way such a shift in group structures can be forced experimentally, as we have done in the

previous experiment, is via assigning to the participants the task-level objective of describing only the Tangrams which they can see, so as to force a controllable imbalance in their levels of involvement in identifiable segments of the conversation. It should be emphasised that the results both in this chapter and the previous one, as regards the disparity between the salience of antecedents for peripheral and primary participants, are independent of topic (all the elliptical fragments and their antecedents are separated by talk about one Tangram exactly). The claim, in other words, is that peripheral and primary participants do not have equal access to the local context, which is constituted by a set of utterances ‘about’ the current Tangram being described, hence within descriptions of the current Tangram.

Arguably, the direct counterpart of this situation in normal everyday conversation - where making a reference precisely may not be the point - is topic shift, whereupon a third participant suddenly becomes a peripheral participant or an overhearer semantically, e.g. A, B and C are talking about a meeting they had yesterday; A and B then move on to discuss a joint project to which C is not privy; C has at this point changed status. What we have argued is that if it is indeed the case that C does not actively participate within the current topic (say because of considerations of politeness) then her context will diverge from A’s and B’s.

Moreover, none of this is to claim that topic has no role to play in how contexts are distinguished. It certainly does (e.g. within Ginzburg’s model of context summarised in section 2.7.1, a topic, modelled as a question under discussion, *is* the context or at least its most important feature): in dyadic conversation, going back to a prior topic requires the extra effort of reinstating that topic, via re-raising of the relevant parts of the utterance in that prior segment to which one wants to return and this is also often marked (by cues such as ‘anyway’). So, topic shifts certainly affect whether elliptical utterances or anaphora are licensed, and this is certainly worth a separate study. But in the current studies, these effects have been factored out. Thus, if contexts can be said to be indexed to or distinguishable by topic, then our claim is that they’re also indexed to the participants.

7.2 Method

7.2.1 Task and Materials

Basically, in the experiments carried out here, we use the same task and materials as those in the previous experiment (henceforth the Same-Different experiment). Namely, we use the online 3-person variant of the Tangram task with 2 Directors and 1 Matcher; D1, D2 and M. The Directors

proceed through the sequence of 12 figures one at a time by clicking, as appropriate, in the window (figures 7.1 and 7.2). They describe the items one by one to the Matcher so that she, in turn, can identify them and drag and drop them into the appropriate slot. The Matcher's interface is identical in this experiment to the Same-Different experiment.

In the previous Tangram experiment, for every Tangram/item it was either D1 or D2 - never both - which could see and were responsible for describing the Tangram to the M; i.e. we gathered a sequence of *dyadic* interactions between D1 and M and some between D2 and M.

The crucial difference - and this is the only difference - between this experiment and the Same-Different experiment is that some of the items are *collective/collaborative*; i.e. the item is *shared* between the Directors such that they can both see it and are told that for these items they are to collaborate in describing the item to the M. The collective items, are tagged with the word 'shared' below the item number (see figures 7.1 and 7.2), so that before one of the directors clicks the item to reveal/turn it over, they both know that the item is a collaborative one. The first Director to click the item to turn it over will also, automatically, reveal the card for the other Director.



Figure 7.1: Director 1's view

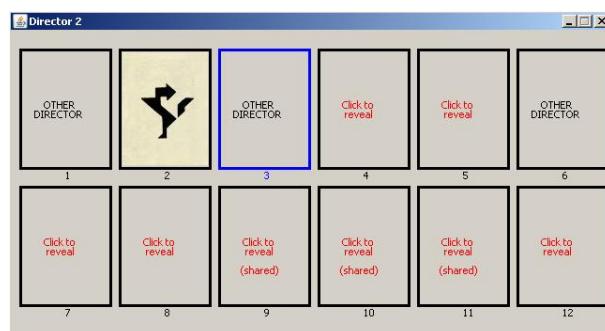


Figure 7.2: Director 2's view

In this manner, for each set of 12 Tangrams, we gather a sequence of interaction between

D1 and M, D2 and M, and unlike the Same-Different experiment, some between all 3, with the sequence controlled experimentally.

In order to enable the insertion of probe clarification questions, all interactions between the Directors and Matcher pass through the DiET Server (described in section 5.2.2).

7.2.2 Probe Reprise Fragments

In this experiment, as in the Same-Different experiment, we parse the Directors' utterances in real-time on the DiET server, in order to identify target antecedents for probe RFs (Reprise Fragments). The criteria used to identify these were identical to those in the Same-Different experiment, viz. that they should be a noun phrase, verb phrase or prepositional phrase, that is not subsequently repeated prior to its use in a spoof RF (in order to factor out the element of ambiguity with respect to the current figure). Given a suitable target, the server can then generate a spoof clarification question that appears to originate from the Matcher, but is not relayed to the Matcher's screen. Both directors see the RF and either may respond. The response is trapped by the server (such that the Matcher does not see the response) and the dialogue resumes.

Note that since the criteria for choosing an utterance as the target antecedent for the RFs were identical to the Same-Different experiment, we can compare our results here with those of the previous one. In other words, we can assume safely that any difference between the patterns observed here and those in the Same-Different experiment are not an artefact of the RFs per se.

The probe RF interventions are, as before, triggered when the Matcher places a figure in the appropriate slot, and just before everybody moves on to discuss the next figure. The target antecedent of the RF is always from the previous item, i.e. the fragment queries something said previously by one of the directors about the penultimate figure described so far.

7.2.3 Experimental Design

As noted, in the experiment reported here the probe spoof RF is always generated on an antecedent (target) expression that has been used to describe the *previous* Tangram figure in the sequence. This created a situation in which probe RF's occurred, on average, at equivalent distances from their targets in terms of time and turns elapsed. This allows us to factor out the effect of distance in number of turn or time elapsed.

This experiment uses a *factorial design* with 2 factors, each with 2 levels/values:

- Antecedent Context: whether the distant context/item, from which the target antecedent

RF is chosen (corresponding to the penultimate figure described so far) was:

- *Collective*
- *Dyadic*
- Intervening Context: whether the intervening context (corresponding to the current, local item intervening between antecedent and RF) was:
 - *Collective*
 - *Dyadic*

This results in 4 possible conditions; viz. *Collective Antecedent Context/Collective Intervening Context* (hereafter C/C), *Collective Antecedent Context/Dyadic Intervening Context* (hereafter C/D), *Dyadic Antecedent Context/Collective Intervening Context* hereafter D/C) and *Dyadic Antecedent Context/Dyadic Intervening Context* (hereafter D/D).

Exactly 4 RF's (1 in each condition) are sent to the directors per round of 12 Tangrams, resulting in a *within-subjects* manipulation.

7.2.4 Participants

96 participants (32 triads) were recruited from among undergraduate and postgraduate students from various disciplines at Queen Mary, University of London. The experiments took approximately 60 minutes to complete and each was paid 7 pounds for taking part.

7.2.5 Procedure

Participants were randomly assigned to the Director and Matcher roles. They were given written and verbal instructions about the task together and had the opportunity to ask questions about the procedure. They were then seated in separate rooms, each at a desk with a PC displaying the appropriate task window (Figure 7.2 for the Directors and Figure 6.2 for the Matcher) and the chat client window (Figure 6.3). They did as many rounds of tangrams that they could in 1 hour and were then debriefed about the experimental manipulations together.

7.3 Results

As highlighted in the introduction, the principal question for this analysis is whether collective intervening contexts would have an effect on the accessibility of the target antecedent, when

compared to the conditions in which the intervening context is dyadic. More specifically, we need to contrast the level of attachment to the local context, of the locally peripheral participants, that is, those participants who are non-members in the locally active, interacting parties, with that of primary participants, i.e. those who for the local stretch of talk are members of the interacting parties.

Seen in the light of the results from the Same-Different experiment, we would predict that a distant antecedent, regardless of whether it was produced in a collective or dyadic context, should be more accessible when the owner of the antecedent has not participated in the intervening context, i.e. that antecedents should be more accessible - and hence the RF more likely to be interpreted as distant - when the local, intervening context is dyadic, namely in the C/D and D/D conditions.

Like the Same-Different experiment, the RFs didn't contain any address cues and they were sent to both directors, so either could respond. When a response was given by either director, following the same protocol as in the Same-Different experiment, it was classified as to whether it displayed a *Local* (i.e. incorrect) or *Distant* (correct) interpretation, or no interpretation at all (*None*, when the response itself was a request for clarification).

7.3.1 Effects of Antecedent and Intervening Context on who responds

The owner of the target antecedent answers the probe RF in 43% of cases. The other, non-owner director provides an answer to only 21% of cases. However, there is relatively little overlap. There are only 12 cases in which both provide an answer to the same RF (i.e. there are two interpretations, one by each director). Taking this into account, overall, 60% of the probe questions receive answers from at least one of the directors.

A Multinomial Regression analysis was performed to test for whether Intervening Context (Collective vs. Dyadic) or Antecedent Context (Collective vs. Dyadic) had an effect on who responds. This showed no reliable effects for either factor (Intervening context: $\text{Chi}_{(1)}^2 = 0.07$, $p(\text{two-tailed}) = 0.79$; (Antecedent Context: $\text{Chi}_{(1)}^2 = 0.48$, $p(\text{two-tailed}) = 0.48$).

As noted however, the antecedent owner was much more likely to respond than the other director (Figure 7.3). This was despite the fact that some of the antecedents were produced on collective items.

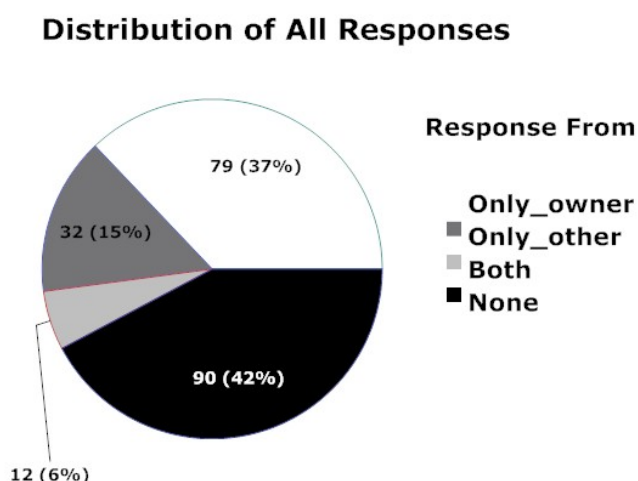


Figure 7.3: Distribution of All Responses

7.3.2 Effects of Antecedent and Intervening Context on Accessibility of Antecedents

To test whether the context of the most recent item (Collective or Dyadic) or the context of the target antecedent (Collective or Dyadic) had an effect on the directors' interpretations of the RFs (local, distant or none), a Multinomial regression analysis was performed on interpretations with two factors (Intervening Context and Antecedent Context). This showed no main effect of Intervening Context ($\text{Chi}_{(2)}^2 = 1.58$, $p(\text{two-tailed}) = 0.45$) and no main effect of Antecedent Context ($\text{Chi}_{(2)}^2 = 1.49$, $p(\text{two-tailed}) = 0.47$), on interpretations.

A second Multinomial regression was performed on Owner responses only to check that the failure to find any effect was not due to noise added by the non-owner's responses. This showed the same pattern with no main effect of Intervening Context ($\text{Chi}_{(2)}^2 = 2.13$, $p(\text{two-tailed}) = 0.34$) and no main effect of Antecedent Context ($\text{Chi}_{(2)}^2 = 0.27$, $p(\text{two-tailed}) = 0.87$).

In order to test whether Owners produced a different distribution of responses to Non-owners, a multinomial regression analysis was carried out with the source of the answer (Owner vs. Non-owner) as a single factor and response types (Local, Distant or None) as the dependent variables. This showed a significant main effect of source ($\text{Chi}_{(2)}^2 = 13.16$, $p(\text{two-tailed}) = 0.00$). Non-owners were less likely to provide a distant interpretation ($B = -1.83$, $p = 0.00$) than owners but there was no difference in their likelihood of finding a local interpretation ($B = -0.65$, $p = 0.21$). Overall, this shows that Owners are more likely to correctly identify the antecedent.

Overall the results show no simple effect of Intervening Context nor of Antecedent Context on the accessibility of the target antecedent (but see below). The owners are more likely to correctly identify the antecedent and more likely to respond but this is not affected by either of

	Local	Distant
Owner Answer	37	50
Non-Owner Answer	20	8

Table 7.1: Distribution of Interpretations by Owners and Non-owners

the context factors.

7.3.3 Activity Levels of Participants over Collective Vs. Dyadic Items

To test whether there were different patterns of contributions for the collective or dyadic items an Analysis of Variance was performed with Items (Collective vs Dyadic) as a between-subjects factor and average number of words, turns and time taken as dependent measures. This showed no effect of Item on average number of turns ($F_{(1,61)} = 9.4$, $p = 0.23$: Collective Mean Turns = 13.18, $SD = 6.9$; Dyadic Mean Turns = 12.4, $SD = 5.9$), no effect on average number of words ($F_{(1,61)} = 0.02$, $p = 0.88$: Collective Mean Words = 51.9, $SD = 29.6$; Dyadic Mean Words = 53.2, $SD = 32.9$), and no effect on average time take ($F_{(1,61)} = 2.40$, $p = 0.13$: Collective Mean Time = 84.8 seconds, $SD = 42.6$; Dyadic Mean Time = 103.5 seconds, $SD = 52.8$).

This analysis however, takes all participants' contributions (including the matcher) into account. In order to see the distribution of the directors' contributions over dyadic vs. collective items, the average number of Turns and Words produced by each of the directors for each Item was entered in an Analysis of Variance with Item Type (Collective vs. Dyadic) as a between-subjects factor. This showed an effect of Item Type on Turns ($F_{(1,157)} = 4.55$, $p = 0.03$) with more turns on average for Collective Items (mean = 4.7, $SD = 2.8$) than Dyadic Items (mean = 3.9, $SD = 2.2$). The parallel analysis for number of Words produced by each director however, showed no effect of Item Type ($F_{(1,157)} = 0.14$, $p = 0.71$). Mean number of words for Collective Items was 19.0 ($SD = 11.0$), and Mean number of words for Dyadic Items was 18.3 ($SD = 11.3$).

Below, in the discussion section, we will use the evidence here, to argue that the directors act as a party on collective items.

7.3.4 Comparison with the Same-Different Experiment

The fact that there was no main effect of Intervening Context (Collective vs. Dyadic) on interpretations, casts a little doubt on the results that we obtained in the previous chapter. We had predicted that Dyadic Intervening Contexts should make the antecedents more accessible and hence make distant interpretations more likely. A subset of the data was therefore inspected, to

see if this experiment (henceforth the Collective-Dyadic experiment) replicated the results of the Same-Different experiment:

In order to test whether the Collective-Dyadic experiment replicated the results from the Same-Different experiment, we looked at the subset of conditions in the Collective-Dyadic experiment that reproduced conditions in the Same-Different experiment. There were two cases. The most directly comparable cases in the present experiment were those in which the Antecedent Context and the Intervening Context were both dyadic. In this case there is a direct parallel with the previous experiment in terms of whether the person producing the target antecedent for the RF also describes the intervening item (Dyadic-Same) or not (Dyadic-Different). In addition, for cases in which the Antecedent was collective but the Intervening item was dyadic we could also code whether the person who actually produced the target antecedent of the RF also described the intervening dyadic item. This gave us a total of 31 cases, 11 involving a switch and 20 involving no switch of speaker between Antecedent and Intervening item.

From this analysis, two cases were excluded because they involved a sub-dialogue between the matcher and the non-owner of the intervening item, immediately prior to the probe RF, about another, unrelated, previous item (not the item from which the RF was taken).¹

A Chi square analysis on the frequency of Local and Distant interpretations in these recoded Dyadic-Same / Dyadic-Different cases showed a reliable effect ($Chi^2_{(1)} = 4.97$, $p(\text{two-tailed}) = 0.03$, Fisher's Exact Test: $p(\text{two-tailed}) = 0.03$). This showed that the present results are consistent with the Same-Different experiment.

The lack of effect of the Intervening Context on the antecedent owners' interpretations seems therefore to be due to the antecedent owner, also describing the local, current intervening item (the Dyadic-Same cases, corresponding to the 'same dialogue context' condition in the Same-Different experiment, in which we would predict high frequencies of local interpretations). These comprised nearly half of our data points in the Dyadic Intervening Context condition, and hence caused the lack of effect of this factor on interpretations. As we will see below, if we exclude these cases from the Dyadic Intervening Context condition, in the contrast between Collective vs. Dyadic Intervening Context conditions we get a strong indication of the higher accessibility, for

¹Recall in this regard, that whether contexts are dyadic or collective, was coded according to whether both directors could see an item or not. In these two cases, the item was indeed dyadic in this sense. However, in both cases, the director who is supposedly an inactive, peripheral participant because she couldn't see the intervening item, was in fact locally active/involved, because she was having a dyadic conversation with the Matcher about some previous item mistakenly placed. Note that we inspected the stretches of talk immediately prior to *all* the probe RFs involved in this analysis to exclude such cases.

the antecedent owners, of the antecedents in the reduced Dyadic Intervening Context condition than in the Collective Intervening Context condition.

	Local	Distant
Same Participant	8	12
Different Participant	0	9

Table 7.2: Distribution of Interpretations when the Owner of Intervening Item is the Same as or Different from the Antecedent Owner

Nevertheless, there are rather puzzling differences between the Collective-Dyadic and the Same-Different experiments. In the Collective-Dyadic experiment, out of all possible responses from the directors (Number of Probes * 2), 32% were actually produced by the directors. In the Same-Different experiment, of all possible responses, 34% were produced. Table 7.3 shows the distribution of all response types received in the two experiments.

As one might intuitively expect, there should be significantly more responses in the Collective-Dyadic than the Same-Different experiment, as some of the items were collective. This has not come about.

Moreover, the distribution of response types indicates that in the Collective-Dyadic experiment the directors are more likely than in the Same-Different experiment to identify the antecedents correctly. But since some of the items are collective and hence jointly described by the directors, we'd expect the opposite. Directors' collaboration on such items, should, overall, cause the directors to be more attached to the local context and hence reduce the overall accessibility of the distant antecedents.

The locally peripheral participants - over dyadic items - do not seem to behave differently in the two experiments. They almost never interpret the fragments locally. These correspond roughly to the 'different dialogue context' condition in the Same-Different experiment and to the re-coded Dyadic-Different cases above, in the Collective-Dyadic experiment.

This difference in the relative frequency of distant interpretations across the two experiments, is due to the locally primary participants, being generally more able in the Collective-Dyadic experiment to identify the antecedents. Note that this is not because some of the antecedents are collective, tempting one to explain this by saying that two directors who have collaborated over an item are more likely to identify an antecedent than one director alone. If we consider the antecedent owner's responses alone, the difference between the two experiments still persists: 50% of all the responses received from the owner displayed distant interpretations in the Collective-

	Local	Distant	None
Same-Different	46/90 (51%)	26/90 (29%)	18/90 (20%)
Collective-Dyadic	52/135 (39%)	54/135 (40%)	29/135 (21%)

Table 7.3: Distribution of Local and Distant interpretations in All the responses received in the 2 experiments

Dyadic experiment, while in the Same-Different experiment, only 30% of owner responses were distant.

These differences seem to be important and indeed puzzling, and do cast some doubt on the account that we would like to have been directly supported, which is given below in the discussion section. Unfortunately, we have not been able to find an explanation for these differences.

It is notable that these differences don't appear to be explainable by different patterns of involvement and contribution of the participants in the two experiments. If anything, it might be the case that in the Collective-Dyadic experiment the participants were overall more involved, that is, the non-owners of the dyadic items in the Collective-Dyadic experiment were more involved than those of the Same-Different experiment. This would only make things more puzzling. It would further support the intuition that because of the higher levels of involvement in the Collective-Dyadic than the Same-Different experiment, the directors should be more strongly attached to the local context (because of active participation), and hence that the distant antecedents should be *less* accessible overall in the Collective-Dyadic experiment, not more. We have, nevertheless, carried out a number of analyses comparing the patterns of contribution of the participants in the two experiments.

Patterns of Contribution in the Collective-Dyadic vs. Same-Different experiments

In both experiments participants had 1 hour to complete as many trials as possible.

The average number of turns, number of words and time taken for each item, by each triad, in the Same-Different experiment and the Collective-Dyadic were combined into an Analysis of Variance with Experiment (Same-Different vs. Collective-Dyadic) as a between-subjects factor. This showed neither a reliable difference for Mean number of Turns per Item ($F_{(1,46)} = 0.23$, $p = 0.63$, Mean Turns per Item Same-Different = 11.9, SD = 6.8, Mean Turns per Item Collective-Dyadic = 12.8, SD = 5.7), nor Mean number of Words per Item ($F_{(1,46)} = 0.05$, $p = 0.83$, Mean Words per Item Same-Different = 52.5, SD = 24.4, Mean Words per Item Collective-Dyadic = 54.4, SD = 32.6), nor Mean Time per Item ($F_{(1,46)} = 1.71$, $p = 0.19$, Mean Time per Item Same-Different = 115.9 seconds, SD = 50.4, Mean Time per Item Collective-Dyadic = 99.6, SD =

50.8).

To test whether the number of items completed by triads was different in the two experiments the total number of items per triad was compared with experiment as a between-subjects factor. This showed no reliable difference (Mann Whitney $U = 249.0$, p (two-tailed) = 0.87). In the Collective-Dyadic Experiment triads completed an average of 25.8 items (SD = 12.1) in the Same-Different Experiment the average was 26.2 (SD = 6.9).

So, overall there is no difference in the average number of turns or words used per item by directors or matchers in the two experiments. There is, however, a difference in the patterns of contribution.

In the Same-Different experiment the number of directors' turns were negatively correlated i.e. the more one director contributes on an item the less the other does (Kendalls Tau = -0.153, $p = 0.00$). This is expected since all the items are dyadic. In the Collective-Dyadic experiment, the number of director turns is positively correlated (Kendalls Tau = - 0.174, $p = 0.00$), so the more one contributes on an item, the more the other does.

However, the positive correlation between the number of directors' turns in the Collective-Dyadic experiment, seems to be solely due to the their collaborations over collective items and not dyadic ones, since the non-owner directors of the dyadic items in this experiment, don't appear to behave differently in this experiment and the last. In both cases, they hardly contribute anything:

To test whether the non-owners contribute more during discussions of Dyadic items in the Collective-Dyadic experiment than in the Same-Different experiment, an Analysis of variance with Experiment as a between subjects factor was carried out on the average number of words and turns produced by non-owners. This showed no reliable difference for the Mean number of Turns per item ($F_{(1,47)} = 0.43$, $p = 0.51$ (Collective-Dyadic Mean Turns = 2.13, SD = 1.6, Same-Different Mean Turns = 1.78, SD = 2.1), or Mean Words per item ($F_{(1,47)} = 0.23$, $p = 0.63$ (Collective-Dyadic Mean Words = 8.6, SD = 10.1, Same-Different Mean Words = 7.1 , SD = 10.9)). This shows that in both experiments, on dyadic items, the directors who could not see the item (the non-owners) hardly contribute anything.

7.3.5 Collective Vs. Dyadic-Different Intervening Contexts

As noted above, the lack of effect of Intervening Context on interpretations seems be due to the Dyadic-Same cases described above. That is, those cases in which the antecedent owner also de-

scribed the intervening, local item. Given the results from the previous chapter, we would expect a higher frequency of local interpretations in such cases, because of the stronger attachment of the antecedent owner to the local context.

If we exclude these cases from the Dyadic Intervening Context condition, and include only the Dyadic-Different cases, we get a strong indication of the higher accessibility, for the antecedent owners, of the antecedents in the systematically reduced Dyadic Intervening Contexts than in Collective Intervening Contexts. In particular, the RFs in the Collective Intervening Context condition seem equally likely to get a distant or local interpretation from the antecedent owners, whereas in the Dyadic-Different cases, they almost never interpret the fragment locally (Figure 7.4).

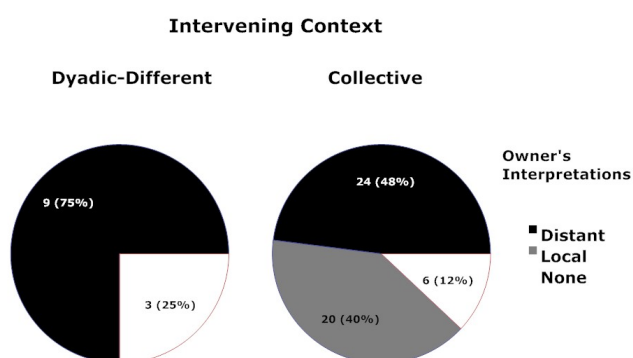


Figure 7.4: Interpretations: Collective Vs. 'Dyadic-Different' Intervening Context

7.4 Discussion

7.4.1 The directors act as a *party* on collective items

As noted in the introduction, the results from chapter 5 suggest that when 2 or more participants have similar responsibilities, as publicly presumed by all the ratified participants, with respect to the current topic of the conversation, they are organised into a party, such that the party acts as a unified aggregate to carry the conversation forward such that one party member can stand proxy for others in doing this.

On this basis, it seems that the 2 directors should act as a party in describing the shared, collective items to the matcher, such that the primary axis of the conversation during these collective items is between the 2 directors as a whole and the matcher rather than between the three individuals. By contrast, on dyadic items the director who cannot see the item, and thus does not have any concern/responsibility as to what goes on between the Matcher and the other Director,

falls outside the interacting parties (single individuals in this case).

What evidence is there that the 2 directors are acting as a party on collective items in the manner explicated in chapter 5? And moreover, what evidence is there that such party formations are tied to topic, such that upon topic shifts new parties may be established and old ones suspended? As we will argue below, there's strong evidence that the 2 directors act as a party on collective items.

- There was no difference between the average number of words produced by (both of) the directors over dyadic vs. collective items.
- The comparison between the patterns of contribution between this experiment and the last shows that in both, on dyadic items, the directors who do not see the figure, hardly contribute anything.
- There was a very small but reliable difference between the average number of turns produced by each director in dyadic vs. collective items (3.9 vs. 4.7).

Taken together, these suggest that the effort put in by one director alone to describe a dyadic item, becomes, on average, equally shared between the two directors on collective items. That the number of turns produced by them is slightly higher on collective items, strongly suggests that, on average, corresponding to each longer turn produced by one director over a collective item, if there's another turn by the other director, it would be proportionally shorter. This shorter turn might be an acknowledgement of the other's description, or one that is highly elliptical with the other's turn as antecedent. This should happen quite rarely though, as otherwise we'd expect the number of directors' turns on collective items to be nearly twice as many as those on dyadic ones. It therefore seems that the directors often follow-up on one another's descriptions, and tend not to repeat the same description. There's little redundancy of descriptions between the two directors on collective items.

This means that, all things being equal, a description by one director can stand proxy for the other's description, and that the primary axis of the dialogue is between the 2 directors as a whole and the matcher. Thus, on collective items, they seem to be acting as a party in the sense explicated in chapter 5, whereas on dyadic items, one director becomes an inactive peripheral participant, i.e. a non-member in the interacting parties.

That the directors act as a party on collective items is, in addition to the above, evidenced by the patterns of Collaborative Completions or Split Utterances produced by the participants in this

experiment. Recall that one of the ways in which parties may be projected, or party membership displayed is according to Lerner (1993), via collaborative completions or split utterances (Lerner, 1991), by a participant of another's utterance such that both parts of the completed turn are addressed to someone else. In this manner, the 2 participants display to a third that they are in alignment with respect to the current topic; and in the absence of negative feedback from the first participant on the completion by the second, they are equally responsible/committed to what was jointly uttered. Such split utterances also seem to be one way in which parties already projected and relevant may become manifest in interaction. In short, split utterances of this sort, provide an empirical measure of the degree to which participants see themselves and act as members of a party, in the sense explicated in chapter 5.

We therefore also carried out a post-hoc analysis of some of the conversational transcripts gathered in this experiment, for split utterances according to the protocol developed by (Purver, Howes, Healey, & Gregoromichelaki, 2009, to appear). This protocol has been reliably applied between independent coders. In addition, the split utterances were coded according to who produced the two parts of the split:

- Within (party): the two directors
- Across (parties): one director and the matcher
- Self: the same participant

As table 7.4 indicates, splits between the directors' utterances (Within) appear almost exclusively on collective items. Moreover, although explicit address cues are rare in the dialogues, it is implicit that both parts of these Within splits, produced by the directors are addressed to the matcher (as the director who is the speaker of the first half of the split never responds to the second director's completion. The matcher does). This indicates, according to what we discussed above, that the 2 directors act and see themselves as a party only during collective items, while on dyadic ones, one of them falls outside all the interacting parties (single individuals in this case).

Table 7.4: Distribution of Split Utterances Over Collective vs. Dyadic Items

	Within	Across	Self	No Split
Collective Items	24 (3%)	25 (3%)	121 (14%)	672 (80%)
Dyadic Items	1 (0%)	62 (4%)	260 (17%)	1212 (79%)

Taken together, the above constitute strong evidence that the directors are acting as a party in the sense explicated in chapter 5, but only on collective items. This therefore shows, in addition, that party formations are tied to topic, such that upon and in virtue of topic shifts, the arrangements of the participants into the active, interacting parties could change.

7.4.2 Sensitivity to who said what

We pointed out above, that regardless of whether the antecedent context was collective or dyadic, the antecedent owner is more likely to respond to an RF than the other director. And when the other director does respond, she often responds with a local interpretation and not a distant one. Whenever she does respond with a distant interpretation, except in 1 case, the antecedent context is collective - as would be expected, since in dyadic antecedent contexts the other director bears little or no responsibility whatsoever towards what is being said as she does not see the item.

The paucity of distant interpretations in the non-antecedent-owner directors' responses when compared to the antecedent owners', suggests that they have not been able to interpret the RF relative to the distant antecedent context, even when it was collective. So even when some participants are organised into parties, a party member, is more sensitive to what she, herself, has said. She remembers it better than what other co-members have said.

But anecdotally, in normal circumstances, such fragments would have more local antecedents and therefore be easily interpretable for all the party members. Given the higher likelihood of responses from the antecedent owners, the responsibility to repair/discuss a previous utterance seems to remain, at least initially², with the speaker of that utterance and not the other party members.

7.4.3 New Parties Established: Divergence of dialogue contexts

As just discussed, there's strong evidence that the directors act as a party on collective items. The analysis of the accessibility of antecedents (distribution of local vs. distant interpretations), however, is not fully conclusive. Nonetheless, it seems to be strongly consistent with the account that follows:

The distribution of the antecedent owners' interpretations of the RFs, when the intervening context is collective, indicates that when a participant is locally a member of the interacting parties, and hence is a primary participant, there's a strong contention between the contexts

²'at least initially' because we showed in chapter 5 that failure by one member to address an issue, will lead to some sort of compensation, by the other members

against which, a RF with a distant antecedent, sent to them at that point, is assessed. There's a considerable tendency on the part of primary participants (party members) in this experiment to interpret the fragment not against this distant context, but the local one (this tendency, however, is reduced from the previous experiment). As mentioned above, the antecedent owner seems equally likely to get a local or distant interpretation, when the intervening context is collective. This is despite the fact that all these fragments have their antecedents in a prior, distant context and have not been repeated since, and hence contain often unambiguous semantic cues as to the context against which they are to be assessed, indicating that the local context has priority for the active, party members.

By contrast, when the antecedent owner is a peripheral participant with respect to the current Tangam (topic), this contention seems to disappear almost completely. The most striking generalisation across the two experiments, is that the local conversational context is a lot less available to a participant who is locally uninvolved, as compared with the party members who are actively contributing, i.e. the peripheral participants are only weakly attached to the local context. This, as briefly noted in the previous chapter, seems to be because of the publicly held assumption among the ratified participants, that the peripheral participant has no concern, goal or responsibility with respect to the current topic/context: "Why would anyone ask me about the current Tangram? They know I don't see it and I haven't said anything about it". Peripheral participants, then, are more likely to interpret an elliptical utterance against the last context in which they were a primary participant. Here, this is the distant context in which they produced the antecedent.

Peripheral participants then, as those who, in virtue of a new topic, fall outside the new parties (comprised of single individuals in this experiment) projected or made relevant by that topic, are distinguished from the members of these parties, in that the dialogue context is 'suspended' for the peripheral participants, while it 'moves forward' for the members of the locally interacting parties. The local context, that is, the set of utterances being currently exchanged in the interaction between the current parties is a lot less available to the peripheral participant as antecedents of elliptical utterances. For such peripheral participants, the (distant) questions or topics with respect to which they were last a primary participant, remain more salient as the background against which elliptical utterances, addressed to them, get resolved. Therefore, at topic junctures where the arrangement of the participants in the active, interacting parties changes, dialogue contexts diverge systematically according to who is and isn't a member of the newly established

parties.

If the account given above is correct, what this experiment adds to what we found in the previous chapter, is that a side-participant party member, independently and in spite of her temporary lack of overt feedback, will be more strongly attached to the local context, than a peripheral side-participant who in virtue of the current topic is not a member of any of the interacting parties and is only weakly attached to the local context. The primary side-participant's view of the dialogue context 'moves forward', evenly, with the speaker or addressee co-member. In the previous experiment on the other hand, the primary participants were always speakers or addressees, and therefore questions as regard the divergence of dialogue contexts for party members and non-members weren't addressed.

In short, the results from this chapter indicate that one and the same utterance can give rise to systematically divergent contextual increments on the information states of party members vs. non-members, resulting in the systematic divergence of the context of the dialogue for these participants, manifested in the disparity between the interpretations assigned to similar RFs received by them.

7.5 Conclusion

In this chapter we provided experimental evidence that party formations are tied to topic, such that at topic junctures new parties may be established and old ones suspended.

The analysis of the accessibility of distant contexts/antecedents, shows that at topic junctures, when a participant is removed in virtue of the new topic from a previously projected party, and thus becomes a peripheral participant, the context that she shared as a primary participant (party member) with other primary participants before the new party was established, is 'suspended'. All the while, the shared context between those who 'carried on' as primary participants for the duration of the new topic 'moves forward'. This is consistent with the results from the previous chapter, in that a locally uninvolved and indifferent peripheral participant is only very weakly attached to the local context (the current topic and the set of sub-topics thereof) and hence very likely to interpret an elliptical utterance addressed to them, against the background of a distant context in which they were last a primary participant.

Particular to this chapter is the experimental evidence, indicating that party members and non-members effect systematically divergent contextual increments/updates. This gives rise to the systematic emergence of distinct dialogue contexts in conversations where the organisation

of participants into parties is, in virtue of topic shifts, dynamic. This is a characteristic which is likely to hold of a considerable proportion of naturally occurring dialogue.

Moreover, the results indicate that the responsibility for (initial) repair and discussion remains with the speaker of an utterance, even when the utterance is produced by a member of a party. In other words, the participant who produces an utterance is much more likely to address any problems in others' understanding or discussion regarding that utterance, even when this participant is a member of a party.

In the next chapter we will explore the likely implications that the patterns observed in this chapter and the previous two could have for formal/computational models of dialogue context. In particular we will discuss how these findings may be construed within Ginzburg's question based approach to tracking contextual change, and present some tentative ideas as to how the model could be transformed to capture these patterns.

Chapter 8

Concluding Discussions: Ramifications for models of dialogue and future work

8.1 Introduction

In this chapter we will first summarise the empirical findings and the overall argument of this thesis. We will then try to explore the likely implications that these findings have for Ginzburg's question based approach to modelling the dynamics of dialogue context. It should be pointed out at the outset that we will not delve into the formal details of Ginzburg's model, and indeed that the details of the choices to be made, of interaction protocols and corresponding contextual update rules in the model are, partly, underdetermined by the empirical phenomena observed here. Nonetheless, these phenomena do seem to create significant challenges for the existing approaches and these will be highlighted.

8.2 Empirical Findings

8.2.1 Parties in conversation

We have shown that multi-party dialogues can lead to collective states of understanding or shared contexts that, unlike those in dyadic dialogue, are not reducible to the individual, pair-wise interactions that gave rise to them. I.e. contexts that are equally shared between and accessible to the participants, in spite of the local/temporary differences in their participant statuses and hence their levels of overt interaction/feedback. We argued that these collective contexts are the result, at least, of those dialogues or segments therein, during which proper subsets of the ratified partic-

ipants are organised into parties. The essence of a party thus formed, is that a dialogue act/move from one of its members is taken to be associated not with that particular individual alone, but with the party as a whole, via a process that we dubbed *grounding by proxy*. In this manner, one member can stand proxy for other members, in performing the action that satisfies the constraints imposed by the current context of the conversation, whatever they may be (e.g. answering a question, signalling acknowledgement/acceptance etc). All the while, the rest of the silent/inactive members are taken to have effected the same contextual updates as the individual member - the proxy - that provided the overt feedback and thus built up the resulting communal/collective context. We provided causal, experimental evidence for the practical reality of parties characterised via grounding by proxy, in the form of local and global compensations on the part of other members for the (apparent) failure of the proxy to ground, i.e. others' compensations for her failure to satisfy the constraints imposed by the current conversational context.

Recall that Schegloff (1995) proposes that the rules of the turn taking system operate *in the first instance* between parties - as higher order entities than individuals - in such a way that turn taking across individuals in dyadic dialogue becomes a special case of turn taking across parties - comprised of single individuals in this case. What we say above, opens up the possibility of a pragmatic equivalent of the same idea. Accordingly, grounding could also operate *in the first instance* between parties rather than individuals. So the contextual updates incurred as a result - or the pragmatic force - of a dialogue move from one party member, are - in the absence of negative feedback from other members - applied to the party as a whole (same for all the rest of the members of that party). As we'll discuss in more detail below, contextual update rules akin to those in Larsson (2002), could remain the same as in dyadic dialogue, but instead operate on/between parties rather than individuals (though see below on what happens when a participant falls outside all active/interacting parties). Grounding in dyadic dialogue then becomes a special case of this where the parties in question are comprised of single individuals.

Parties also become manifest in participants' patterns of non-verbal behaviour

Another important point is that party formations seem also be manifested in how the participants' non-verbal behaviour is organised. For instance, Healey and Battersby (2009) present an example where, arguably, the three persons in interaction are organised into two parties: two of them are instructors and the other is a learner. This example shows how the gaze of the learner is fixed

upon one of the instructors, A, while A ‘confers’¹ with the other instructor, B, regarding what he (they) should tell the learner. After this conference episode, what B says in response to A’s query/elicitation of help from B, is told back to the learner by A. So, arguably, via the organisation of gaze, the floor between the 2 parties is ‘put on hold’ while the conferring takes place or while, as discussed above, grounding by proxy has its effect.

It is generally expectable that within what Kendon (1976) has called an F-formation, the non-verbal behaviour of the participants, viz. mutual orientation, proximity, posture shifts and gaze, would ‘pick out’ or index subgroups of the participants *as* parties, owing to patterns unique to the members of that party possibly in a ‘privileged’ shared-space (say the shared-space that is between A and B but not accessible to C in the same way). This can be thought of, roughly, as temporary F-formations within F-formations.

When do participants negotiate common ground among all the members?

It seems we are now also in a position to explain some of the data gathered by Carletta et al. (2002). As discussed in section 3.4.4, their claim is that small group discussions can be characterised as a series of pair-wise discussions with overhearers, with process loss on the part of the third silent party (the ‘overhearer’) due to material which is merely overheard. Contrary to this claim, as summarised above, we have shown that there are dialogues in which a *side-participant* can, in the absence of any overt feedback, reach the same level of grounding as the dyad directly engaged in interaction. So there’s no (grounding) process loss in this case (but see below about peripheral participants). Support for the claim that small group discussion is primarily comprised of pair-wise discussions comes from Parker’s (1988) data, which we have talked about previously. But we should note that, according to the above, it cannot be assumed that these ‘pair-wise’ discussions are, in fact, dyadic. In other words, via grounding by proxy, it doesn’t really matter who is doing the talking so long as he is acting as a spokesperson for a party, including other members who may be silent for the moment. So this claim is potentially misleading.

Moreover, Carletta et al. (2002) claim that, in small group discussions which require a high degree of mutual understanding at every stage, common ground is negotiated explicitly among all group members. We argued that the design of their experimental task was forcing this. We can, now, say more about this: their experiment involved 2 route-givers and a route-follower. So

¹in the sense used by Lerner (1993) and discussed in chapter 5: upon the failure by one of the party members to ground

the two route-givers, on the view developed in chapter 5, interact with the route-follower as a single party. The fact that the route-givers were explicitly negotiating common ground between themselves, is predicted by our notion of grounding by proxy. As discussed, this predicts that if a member of a party fails to ground a contribution, other members should somehow compensate. We showed in chapter 5, that this compensation often takes the form of ‘conferring’ among the party members. Given the discrepancies between the two route-givers’ maps, we’d expect one route-giver to often fail to ground what the other is telling the follower (e.g. when she uses referents/landmarks absent on the other’s map), leading to direct interactions/conferring among the party members themselves. This data, in fact, provides indirect support for our claim about the effects entailed by grounding by proxy. We might say, in general, that grounding by proxy entails that the less aligned the relevant parts of the information states of party members are to begin with, the more intra-party conferring there will be.

Given all this, one hypothesis is that the ratified participants who fall inside and outside all active/interacting parties for a local stretch of talk, dubbed as *primary* and *peripheral* participants respectively, need to be distinguished in terms of grounding responsibilities - entailed by grounding by proxy - and correspondingly the level of grounding (understanding and acceptance) that, in the absence of overt feedback (i.e. as side-participants), they are assumed to have reached.

8.2.2 Emergence of distinct Dialogue Contexts

Given the above and the additional corpus evidence relating to long-distance resolution possibilities for ellipses discussed in chapter 6, we hypothesised that the locally involved/primary participants should be strongly attached to the local conversational context (expressed in terms of the higher saliency of potential antecedents in this context and a strong tendency to interpret distant fragments incorrectly against them), and should thus have an impaired level of access to a distant context/antecedent. By contrast, the locally peripheral participants should have no such attachment to the local context and therefore the distant context, in which they were a primary participant, should be more salient/accessible for them. We showed experimentally that this is indeed the case: while it was the utterances in the local conversational context (an overarching question and a set of sub-questions during which the current parties remain unchanged) that were available and had priority for the primary participants as antecedents of the (elliptical) utterances received by them, for a peripheral participant, this context was a ‘distant’ set of utterances, during which they were last (at some earlier stage of the dialogue) a primary participant. In other words,

there was a systematic disparity between primary and peripheral participants' interpretations of '*equally elliptical*' (on average) Reprise Fragments, *received from the same individual*, all of which were separated from their antecedents, by, on average, the same number of turns/topics.

We argued that this is evidence for the systematic emergence of distinct dialogue contexts within one and the same conversation, as a result of one or more previously primary participants falling outside the active/interacting parties for a local stretch of talk/topic. As we will argue further below, at least some of the standard utterance by utterance update mechanisms of the context that they shared as primary participants with other primary participants, prior to when the new parties were established, are suspended. All the while, the corresponding parts of the shared context between those who 'carried on' as primary participants 'moves forward' (See below for how this may be construed within Ginzburg's model). Consequently a single utterance can result in systematically divergent contextual updates, applied to primary and peripheral participants' information states/contexts, a phenomenon we dub 'Pragmatic Pluralism' in dialogue.

8.3 Two systems of Participant Roles: Party Membership vs. Goffman's Participation Framework

Given the above, it seems that in conversation, there are two systems of participant roles, both dynamic and emergent from the interaction itself, but with distinct grounds on which they are distinguished.

The first, elaborated at length previously, is Goffman's (1981) participation framework, constituted by a dynamically assigned set of roles. The ratified roles within the participation framework - which are what we've been concerned with throughout this thesis - are determined by reference to who, locally in an utterance-event is addressed. In other words, the speaker either by the design of her utterance, or through gaze, posture, gesture and/or the use of another's proper name, selects them as next speaker. Those others not selected in this fashion are side-participants to that utterance. The ratified roles in the participation framework thus determine who should speak next.

The second system on the other hand, is that of peripheral vs. primary participation. As mentioned above, the difference between a primary and a peripheral participant seems to be one of responsibility/concern/commitment, publicly known and assumed by all the (ratified) participants, with respect to the current conversational topic. These roles can be more persistent, since often, a single overarching topic can be under discussion throughout a whole conversation or a

long segment therein.

These systems of roles aren't totally independent since, as we have discussed in chapter 5, the participation framework can be exploited to project parties; e.g. via collective addressing, making the addressees equally responsible for responding to what has been uttered. Nonetheless, if we assume, as we have done, that a party is established and relevant, then a primary side-participant, we have shown, has an equal status with respect to grounding/contextual updates to a primary addressee (except for the fact that the addressee is responsible to speak next and thus stand proxy for the other (side-participant) members). Moreover, a primary side-participant is treated differently from a peripheral side-participant in terms of contextual updates (we will try to express this below in terms of Ginzburg's QUD-Downdate and FACTS-Incrementation).

The distinction between peripheral and primary participation, thus allows us to abstract away from the local, utterance-level configurations of the participation framework - which seem to affect solely who should speak next - in determining what contextual updates should take place for the participants. Instead of the individual utterance, it is the topic and the relations holding between a participant's publicly presumed responsibilities/concerns with respect to that topic, that determine what kinds of contextual updates should take place for her.

This provides us with a way of clustering these roles where the original (ratified) hearer roles in the participation framework are used to manage turn-taking obligations while party status/membership determines contextual updates for side-participants. Note that some combinations aren't possible: an addressee is never a peripheral participant. So effectively, this just allows us to distinguish peripheral from primary side-participants, which in a way, constitutes the whole of what this thesis is about.

Now recall Clark and Carlson's (1982) distinction between Addressee Directed and Participant Directed speech acts (very briefly discussed in section 3.6). The former correspond to the original speech act types, proposed by Searle (1969) (Request, Promise etc.) and the latter to the corresponding 'informative' acts directed at all the participants. According to this distinction, whereas the illocutionary force of a speech act affects only the addressees, the side participants are only 'informed' of this, and hence only acquire the belief that such and such speech act was performed. We may now say, given the notion of party membership, that the Addressee Directed acts should really be Party Directed Acts. As we have shown, when a party is contextually relevant and already projected, in the absence of negative feedback from a side-participant co-member, the contextual increments and the commitments brought about as a result of a di-

alogue move (speech act) from one party member (the proxy), affect the party as a whole. So the illocutionary force of e.g. a promise from one party member affects the rest of the - possibly side-participant - members as well as the actual speaker of the promise. Participant Directed informative acts on the other hand seem applicable to peripheral participants (see below for how this may be construed within Ginzburg's model as divergence of the peripheral and primary participants' DGB.FACTS).

Note that we have appealed to the notion of membership of active/interacting parties, to distinguish peripheral from primary participants, whilst partly characterising the kinds of mutual grounding responsibilities that hold among the members of a party and the kinds of contextual update or common ground increments that in the absence of feedback should take place for them. However, none of this really answers the question of what exactly constitutes a party. What are the necessary relations that should hold between the information states of a set of participants, and a given topic (modelled as a question under discussion in Ginzburg's framework), for them to be regarded as belonging to a single party with respect to that topic? This is an important open question that has not been addressed in this thesis. We have not, in other words, defined parties, but have assumed rather that they are projected and relevant, and have analysed the effects (rather than reasons/causes) of party formations.

Nevertheless, as we have discussed, parties do seem to be strongly tied to topics/questions. In what specific arrangement now, the ratified participants are organised into parties, given a topic/question, seems to be strongly tied to what relations, if any, are presumed publicly (by all the ratified participants), to hold between each participant's concerns/goals and that topic. These relations may be determined in advance, e.g. in formal decision making meetings where the participants have predetermined roles to play, and in all the experiments carried out here. On the other hand, they could, as we have noted in chapter 5, be emergent from the interaction itself, e.g. when two participants create a coalition by adopting the same point of view with regards to a question under discussion, or as noted above, in specific configurations of the participation framework with respect to an utterance-event, e.g. collective addressing.

Henceforth, we will in our considerations of the likely implications of our findings for models of dialogue, be making the simplifying assumption, mostly true in our experiments, that as a question is raised it is clear to the participants in advance, who is a peripheral and who is a primary participant: we'll assume that questions are indexed by or determine the specific arrangement in which the ratified participants are organised into parties; i.e. that questions 'project'

parties.

8.4 Ginzburg's KoS

Ginzburg (2009) states that the basic adequacy criterion of a theory of meaning for dialogue is:

the ability to characterize for any utterance type the update [in context] that emerges in the aftermath of successful mutual understanding and the full range of possible clarification requests otherwise.

This presupposes the view proposed in Dynamic Semantics (e.g. (Stalnaker, 1978; Lewis, 1979)) and adopted by Ginzburg. Specifically, utterance meaning is conceived of as 'context update *potential*'. In other words, meaning is explicated as the effects or changes that an utterance can potentially bring about in the commonly shared contextual resources that conversationalists are able to track. Conversely, utterances are understood or interpreted against the background of the already existing context. That is, the *content* of an utterance emerges as an instantiation, or 'contextual anchoring' of an abstract construct which is the semantics associated with that particular sequence of words, arising from the Grammar. Meanings can therefore be formalised as functions from context to content.

Contexts or Information States, on this view, are therefore seen, as a (structured) set of resources enabling the interpretation of utterances and altered by such interpretations.

The model should, on this view:

1. Capture the highly constrained set of follow-up contributions that are coherent/felicitous at any given point in conversation.
2. For each utterance type it should specify the change/update that the utterance brings about in context (information states of the participants)
3. facilitate the interpretation/resolution of elliptical utterances: Clarification Ellipsis, Short Answers, VP-ellipsis and the like.

In section 2.7.1 we sketched how these can be achieved for dyadic dialogue, in Ginzburg's model of shared context: the Dialogue Game Board (DGB) structured as having 3 components: LATEST-MOVE, QUD and FACTS. We discussed how each of these components contribute to the satisfaction of the above. The dynamics of how each utterance or move, depending on its type

(e.g. querying vs. assertion) changes the context for the speaker and addressee were discussed. We also looked at how interaction protocols arise from the the composition of what Larsson (2002) has called Conversational (update) Rules. Conversational Rules, are mappings from one DGB state to another. The range of possible next utterance/move types and the updates that they give rise to is captured by Conversational Rules.

In this section we'll point out one additional aspect of this theory that is relevant for our later discussion of what changes may be necessary in the model to handle multi-party dialogue. We will, afterwards, discuss the implications of our empirical findings for Ginzburg's approach, with the above criteria in mind.

8.4.1 Divergence of DGBs

What we haven't discussed before which is crucial for our subsequent discussion, is that the speaker and addressee maintain separate views of the context of the conversation, that is, they maintain separate DGBs, corresponding to the public component of each participant's context. The DGB does not however, exhaust a participant's information state. There are private components too (unpublicised plans/beliefs/intentions etc.)².

Although, in dyadic dialogue, the DGBs of the participants are identical most of the time, they can, on occasion, be divergent which is why they are relativised to individual participants in the first place. This happens because: "the speaker is committed to know what his utterance means, whereas an addressee has no such commitment" (Ginzburg, 2009), and hence occurs when there's a misalignment between the information states of speaker and addressee, necessitating meta-communicative clarification interaction. When the addressee realises that she doesn't know what the speaker intended to convey, before clarification interaction ensues, the QUD components of the participants' DGBs are distinct. The addressee has pushed the content question (paraphrasable roughly as "what did the speaker mean?") on her QUD while the speaker has not yet done so. We will not go into any more detail here. Suffice it to say that in modelling dyadic dialogue, distinctness of DGBs draws solely on meta-communicative interaction.

We mention this in order to emphasise that divergence of DGBs is not a systematic occurrence

²The issue of which parts of a participant's information state are public and which private is still a great source of contention. It relates, partly, to the contention among psycholinguists as to what extent conversational partners take each other into account during production/comprehension; so whether participants behave egocentrically or take the unpublicised part of their interlocutor's information state into account (via 'other modelling'). We have discussed this in chapter 2 and will not expand on it any further here (see also Ginzburg (2009), chapter 4 for a discussion)

in dyadic dialogue. As noted, it arises as a result of mishearing, misunderstanding or other more or less 'random' mismatches between the speaker and hearer's points of view. Importantly it does not arise as a result of standard utterance by utterance DGB update mechanisms.

But, as we will argue below, divergence of DGBs is a systematic effect in multi-party dialogues. In fact, it seems that, at least, conversational update rules responsible for the utterance by utterance transitions of QUD, as tracking the resources available to participants for ellipsis resolution need to be relativised to participant role (As determined by party status/membership, discussed above and not in Goffman's sense which seems to be responsible for tracking turn-taking obligations).

8.4.2 Conversational Update Rules could operate on *parties* in the first instance

It seems that the most important choice to be made with respect to scaling up conversational rules, motivated for dyadic dialogue, is to what extent these rules are to be relativised to participant role. Above, we discussed two systems of participant roles: Goffman's original ratified roles and primary vs. peripheral participants. As noted, this allows us to abstract away from the local, utterance level configurations of the participation framework, in determining what contextual update should be effected by each participant. Primary side-participants can reach the same level of grounding as the co-participant (the proxy) who is either speaking or being addressed. We therefore propose that update rules should operate in the first instance on parties - as higher order entities than individuals. This means that the preconditions and the effects of an update rule will be the same for all the members of a party. This is, in effect, the pragmatic equivalent of Schegloff's (1995) remarks about turn-taking rules, operating between parties rather than individuals. Goffman's original ratified roles on the other hand will be used to manage turn-taking obligations. We will discuss this in a little more detail below.

8.4.3 Pragmatic Pluralism in Dialogue: Relativising update rules to participant role

As noted, given the empirical findings summarised above about how peripheral participants are distinguished from primary ones, it seems that update rules governing QUD downgrade/update and FACTS-incrementation do need to be different for peripheral vs. primary participants. In particular, as we will argue in more detail below, it seems that when a peripheral participant is present during/with respect to a question under discussion and the set of sub-questions thereof, the standard update mechanisms responsible for the transitions of QUD at least, are suspended

for the peripheral participant, something that we've called suspension of shared contexts. This implies that one and the same utterance will effect different updates on the DGBs of primary vs. peripheral participants. The DGBs of peripheral and primary participants can thus be systematically divergent. This leads to 'Pragmatic Pluralism' in dialogue. But note that we're not completely in a muddle, as there does seem to be some systematicity to these divergent updates.

As noted, this kind of divergence is very different from that arising from temporary, random mismatches between the participants' information states in dyadic dialogue, which draws solely on meta-communicative interaction. Idealising assumptions of perfect communication, identical semantic ontologies etc. will ensure that the DGBs of the participants in dyadic dialogue are, at least most of the time, identical. However, the systematic kind of divergence in multi-party dialogues mentioned above cannot be accounted for in such terms. Although we demonstrate this divergence using probe Reprise Fragments, it has little to do with miscommunication. Instead the disparity between the peripheral and primary participants' interpretations of these Reprise Fragments points to the distinctness of the questions which are at that point QUD-Maximal for them, resulting as we've argued from the publicly presumed lack of concern, responsibility and participation on the part of the peripheral participant with respect to the question which is QUD-Maximal for the primary participants. One cannot therefore account for the pragmatic structuring of multi-party dialogues - corresponding to all three of the success criteria mentioned above - without taking this into account.

8.5 Contextual updates for Primary vs. Peripheral participants

8.5.1 The Cautiously Optimistic Grounding Strategy

As sketched in chapter 2, according to Clark and Schaefer (1989), most models of dialogue before them, had taken an autonomous view on communication, i.e. they tacitly assumed either that:

- Once something is uttered it is automatically added to the common ground.
- Or once something is uttered it is automatically added to the common ground unless there's evidence to the contrary (negative feedback/repair initiation).

They argued, however, that participants necessarily always collaborate before they can reach the grounding criterion. That is, *positive* evidence of understanding regarding an utterance -

whether tacit in the form of a relevant next turn or explicit in the form of an acknowledgement - should be required before the content of the utterance can be added to the common ground. They present systematic procedures for doing this. These were later spelt out formally by Traum (1994), in terms of Collaborative Conversational Acts.

Larsson (2002) suggests that the difference between the tacit grounding assumptions made by different models is one of degree of optimism. Namely, Clark's account is *pessimistic* in that in the absence of positive feedback it never assumes that something is grounded. Other models criticised by Clark and Schaefer (1989) are *optimistic*: they don't require any feedback and add the content of an utterance to the common ground as soon as it has been uttered.

Larsson (2002) introduces a third strategy for grounding: the *cautiously optimistic*. This is, basically, the same as the optimistic grounding assumption, except that it facilitates backtracking. So, on this view, by default, a participant can assume an utterance as grounded as soon as it's been uttered, but if there's evidence to the contrary (negative feedback), she should be able to revise the common ground, easily and, crucially, without recourse to intricate belief revision mechanisms, so as to reverse all the effects of the erroneous assumption that the utterance was grounded. Larsson (2002) shows that this can be achieved in the model, by keeping the previous state of the DGB 'around', so as to enable effortless backtracking.

8.5.2 Primary Co-participants and an essentially defeasible Communal Acceptance

In multi-party dialogues, this strategy can indeed be very useful. As we showed in chapter 4, side-participants can, in withholding of negative feedback, reach a level of understanding comparable to direct addressees. This makes the cautiously optimistic strategy specially suited for side-participants. Kronlid (2008) suggests, in this regard, that the only reasonable grounding strategy for side-participants is the cautiously optimistic.

Along the same lines, as briefly discussed in section 3.11, Ginzburg and Fernandez (2005) propose the Add Side Participants (ASP) principle, whereby the dyadic interaction protocols in the model are transformed so as to handle multi-party interaction. This is, of course, one of the possible transformations which seems particularly relevant here. It enforces *communal acceptance*, where an acceptance of an assertion by the first participant to respond counts as acceptance by all, and hence results in FACTS-Update and QUD-Downdate, for all those who did not respond. This means that under the protocol induced, once an assertion is accepted by one participant, there's no way for others to disagree with the assertion. The authors mention that

the motivation behind this is the lack of a speaker's capacity to monitor everyone for acceptance signals. This seems right, though as we will see below, communal acceptance cannot really be understood without the opportunity, if not the obligation, for those primary participants who did not explicitly accept the assertion initially, to dissent from it at distance after it's been accepted by a co-participant (the proxy).

Nevertheless, given the results from the previous two chapters, the cautiously optimistic strategy seems only applicable to *primary* participants (importantly not peripheral ones as discussed in the next section below), that is, those co-participants in the same party, on behalf of whom the current speaker (the proxy) does the grounding. As elaborated in chapter 5, the proxy can - on behalf of other co-participants in the same party - be seen as either:

1. responding to one outside the party, that is, providing a second pair-part to a first already uttered by somebody outside that party, i.e. *proxy in response* Or
2. producing a first pair-part addressed at somebody outside the party, i.e. *proxy in speaking*

The difference between these is the utterance or local context, giving rise to the constraints to be collectively satisfied, via grounding by proxy, by the next speaker. We will here explore the first of these two cases, with respect to assertions, since it can prove problematic for interaction protocols which enforce *communal acceptance* and hence will not allow non-adjacent clarification interaction or discussion.

Let's call the person who provides the first pair-part FS and the one who provides the response NS and those for whom NS stands proxy, SP's. NS is either chosen directly by FS - when the addressee-set is a singleton - or she self-selects - if there's more than one addressee, or no particular addressee is selected (addressing by attribution (Clark & Carlson, 1982)).

If the notion of *grounding by proxy* is to be respected, it seems necessary that SP should have the opportunity, if not the obligation to butt in if, from her point of view, the response from NS is divergent from what she would have said or if it fails to secure a sufficient response.³ This divergence can occur on at least two levels of grounding (recall section 2.3.3):

³In a situation where it is clear to all that an acceptance move from an individual belongs to a party as a whole, then the commitments brought about by the acceptance move, in the absence of negative feedback from the other members, also belong to the party as a whole. If one of those who have stayed silent over the issue and hence was taken to have accepted it, subsequently complains, it seems reasonable for the proxy to say: "well if you disagreed you should have said so"

- Identification/Understanding: if SP cannot interpret FS's utterance while NS has acknowledged it.
- Acceptance: if SP does not agree with FS's utterance while NS has accepted it or vice versa.

The first of these would lead to non-adjacent clarification interaction between FS and SP and the second to non-adjacent rejection/discussion. If communal acceptance is in place, both of these would require some sort of backtracking at least on *FS's DGB.FACTS*. We have presented corpus examples of both of these cases in chapter 4 (excerpts 4.3 and 4.5).

To accommodate the second of the above cases (divergence on acceptance), Ginzburg and Fernandez (2005) propose Distributed Acceptance, whereby the acceptance of an assertion does not necessarily involve the adjacency of the assertion in question. Corresponding QUD-Downdate and FACTS-Update rules are in this case distinct for the asserter and the audience, such that the asserter does not downdate her QUD and update her FACTS, unless there's an acceptance move from every addressee, whereas an addressee downdates QUD and updates FACTS, as soon as she herself has made an acceptance move. As a result, an addressee's DGB is not identical to the asserter's until everybody has accepted the assertion. But what if the first addressee accepts and the next rejects? Doesn't the asserter need to maintain a separate FACTS for each member of the audience? What about the audience? A single DGB per participant does not allow any variation between what a participant takes the others to have accepted. After all, we do seem to remember after we've had a conversation with other people, if they disagreed on an issue and with which one we agreed. One can come out of a discussion and say: "A and B didn't in the end agree about Q. But personally I think I was with A on that one.". One FACTS per participant does not seem to allow this. Perhaps the propositions on FACTS need to be indexed to specific participants, which is not very different from indexing FACTS itself. This seems to be, as we will discuss summarily at the end of this chapter, motivation for multiple DGBs per participant.

In any case, the Distributed Acceptance protocol seems unrealistic, since often a primary side-participant or an addressee who does not respond overtly, is taken, as we have shown, to have grounded the assertion in question. Insofar as parties are publicly established and relevant, communal acceptance with backtracking seems like a more empirically viable solution.⁴

⁴It might be worth noting that since communal acceptance is here taken to be a default, the growth of DGBs will be non-monotonic.

The crucial point to be taken in this section is that if any dialogue move by an individual is to be taken as belonging to a party (which we have shown can be the case), then it is only fair that the co-participants in the same party should get a reasonable window of opportunity to dissent from that act. But until when? It seems necessary to allow clarification and discussion at longer distances, at least until a new unrelated topic has been raised by one of the participants, at which point the opportunity is lost and the utterances exchanged grounded for all those primary participants who did not provide any feedback. A question should therefore remain accessible somewhere on the DGB, until a new unrelated question/topic has been raised. But not on QUD, as this would allow infelicitous long-distance NSUs:

Non-adjacent discussion/rejection or clarifications seem to be on average less elliptical. For example, an utterance of a highly elliptical “no” or “what?” almost always attaches itself to the immediately prior utterance, whereas in these cases it is not the immediately prior utterance that needs to be rejected/clarified. It seems therefore that in such cases, certain bits of the utterance to be clarified/rejected need to be repeated, or the clarification/rejection be sentential instead of an NSU, so as to pick out unambiguously the (sub-)utterance that needs to be clarified/rejected. E.g. a Reprise Fragment or a Literal Reprise (as noted previously, Purver (2004) shows that in the BNC, the more distant a Clarification Request is from the trouble-source, the less elliptical its form will be) in the case of clarification and the whole non-elliptical negation of the utterance to be rejected in case of rejections.⁵ But QUD is used in Ginzburg’s model, both to constrain the range of semantically relevant follow ups and to resolve NSUs. Thus, it appears that we might need separate structures to play these roles, as in multi-party dialogues, the unresolvedness of a question does not always coincide with the felicity of an elliptical follow-up specific to that question. We will not go into any more detail in this regard but as Kronlid (2008) points out in his concluding discussion, Larsson’s (2002) ISSUES in addition to QUD, seems to be a possible solution, whereby QUD would contain those questions constraining the set of elliptical follow-ups while ISSUES will contain all questions which have not yet been resolved, including those on QUD. Roughly then, utterances that are to be built as ‘pragmatically next’ to the questions on *ISSUES* \ *QUD* need to be sentential.

⁵This is what we’ve called ‘context re-raising’ (section 3.10), but unfortunately due to time and space limitations haven’t really been able to identify what forms these take, or indeed exactly when it is necessary to re-raise

Grounding in 'chunks'?

For *primary* side-participants, then, the cautiously optimistic grounding strategy seems viable. But, as argued, this is only reasonable if, for any assertion in context, those side-participants who cede the floor to a co-participant viz. the proxy, get a reasonable window of opportunity to dissent from/clarify the assertion, while communal acceptance is also in place. This means that there might be a need for some additional structure on a participant's DGB, which would keep track of the provisionally grounded material, that is, the material which a primary side-participant has produced no feed-back on, while a co-member has in her stead. In this manner if there's any subsequent dissent from the assertion on the part of the side-participant, this material can be easily retracted. On the other hand, if the conversation moves on to a new unrelated topic, all the material on that temporary structure is moved to the asserter's permanent FACTS; hence grounding in chunks.

8.5.3 Contextual updates for Peripheral Participants: Systematic Divergence of DGBs

As summarised at the beginning of this chapter, the experimental data from chapters 6 and 7, indicate that in a single multi-party conversation, there's potential for a systematic conflict between a peripheral and a primary participant's interpretations of similar Reprise Fragments received from the same individual, at the same points in the conversation. This was due to the distinctness of the contexts against which these fragments were being assessed. We argued that this is evidence for the emergence of distinct shared-contexts and that these are indexed to specific sets of participants.

There was a tendency on the part of primary participants, to interpret fragments 'incorrectly' against the local conversational context despite the fact that the fragments contained unambiguous semantic cues referring to a distant antecedent: the primary participants were strongly attached to the local context. For locally peripheral/uninvolved participants on the other hand, this tendency almost totally disappears. They are quite weakly attached to the local context. When they respond, they almost always interpret the fragment correctly against the distant context, that is, the last context in which they were a primary participant. Let's make this more explicit:

Suppose *A*, *B* and *C* are discussing question Q_1 and a set of sub-questions thereof, during which all three are primary participants. They are publicly assumed to be responsible for what is being said regarding Q_1 . Now suppose that the discussion moves on to an unrelated question Q_2 during which one of them, say *A*, becomes a peripheral side-participant, who is

publicly known not to have any responsibility or indeed anything to say about Q_2 and she does not do so. The results from these experiments indicate that, while at this point, a Reprise Fragment with Q_1 as antecedent, exchanged between B and C is often assessed/resolved against Q_2 (strong attachment to the local context), those addressed at A by B or C are almost always correctly assessed/resolved against Q_1 , i.e. while Q_2 has priority over Q_1 as QUD-Maximal on the $B.DGB.QUD$ and $C.DGB.QUD$, on $A.DGB.QUD$, Q_1 seems to have priority over Q_2 as QUD-Maximal. Therefore at that point, $A.DGB.QUD$ is distinct from $C.DGB.QUD$ and $B.DGB.QUD$.

But note that A , the peripheral participant with respect to Q_2 , can, intuitively speaking, always access the material exchanged between B and C about Q_2 elliptically and hence it seems reasonable that she should have Q_2 as Maximal on her QUD as well as Q_1 . But on $B.DGB.QUD$ and $C.DGB.QUD$, it seems that Q_1 should have been downdated. So we might say in this regard that the standard QUD-downdate rules are, for peripheral participants, suspended. Moreover, the ordering on A 's QUD is an essentially *partial* one, as there is more than one QUD-Maximal question. We will not go into any more detail in this regard.

What happens to A 's $DGB.FACTS$ one might ask? Is she, as a peripheral participant during Q_2 , adding the material grounded between B and C to her $FACTS$? This choice is indeed underdetermined by the empirical evidence available, and depends on what $FACTS$ is for. We might say though, that since B and C know that A has no concern with respect to Q_2 , they need not assume anything, at least about A 's level of acceptance. The question of whether A has accepted the material overheard doesn't even arise; e.g. if B and C move on to the subject of where the two of them should meet tomorrow, with A as a peripheral participant, the question of whether A has accepted whether B and C should meet in Chiswick seems irrelevant while she is nevertheless ratified and if asked later, would probably know where A and B are going to meet. So if $FACTS$ is used to represent the grounded material on the acceptance level (representing the public commitments and obligations brought about by the illocutionary force of the utterances exchanged: long-term beliefs, obligations etc.) then we should adopt a pessimistic grounding strategy for the peripheral participant. If on the other hand the grounded material is represented on the DGB separately on each of these levels (e.g. using an obligation layer like that of Matheson, Poesio, and Traum (2000) for acceptance) then we might adopt a cautiously optimistic strategy on the identification/understanding level. $FACTS$ -incrementation for peripheral participants, seems therefore to be also slightly different than those for primary participants, at least on the acceptance level. While in the absence of feedback a primary side-participant is

assumed by the speaker and addressee to have accepted the utterances exchanged between them (and effected the commitments brought about by them), they need not assume anything of this sort for a peripheral side-participant.

8.5.4 Multiple DGBs?

Note that what we say above about FACTS-incrementation for peripheral participants implies indirectly that participants should have more than one FACTS on their DGBs, each representing the commitments arising for every other participant present as these do not seem to coincide for peripheral and primary participants. While the primary participants (party members), upon acceptance of *p*, subscribe - jointly and publicly - to the belief that *p*, they need not assume anything for the peripheral participant. That is, although it seems that the primary participants can assume safely, in the absence of negative feedback, that the peripheral participant has grounded the material on the understanding level (such that the primary participants assume that the peripheral participant knows that they have subscribed to *p*), they need not assume anything about the peripheral participant's subscription to *p*. Participants, then, seem to need to have separate FACTS for every other participant, to represent these divergent commitments.

On the other hand, *insofar as participants keep track of the uneven updates/downdates to QUD*, discussed above, this is further motivation for multiple DGBs on the information state of each participant, one for every other participant present. While an update may take place on one of these DGBs, it need not on the others or rather it need not be the same update.

Nonetheless, we should acknowledge that although our experiments in the previous two chapters clearly show, as formulated above, that within one multi-party conversation, the context shared by a subgroup of participants may be distinct from those shared by other subgroups taken from the same pool of ratified participants, *we have not shown directly* that a single participant keeps track of these, i.e. although we've presented situations where there's a systematic disparity between the contexts against which an elliptical clarification request (a Reprise Fragment) is assessed by different participants (the recipients of the fragment), we have not shown that the participant producing the clarification request is sensitive to this (prospective) disparity. Our experiments do not show, in other words, that a fragment receives divergent interpretations from a single participant, contingent upon whether the fragment comes from a peripheral or a primary participant. Nevertheless, we would very much expect this to be true as it is supported by corpus evidence from Walford (used in chapter 6 to motivate our hypothesis) and as alluded to previ-

ously, indirectly by the evidence in Healey and Mills (2006). They showed that the same probe question receives systematically different answers depending on whether the apparent origin of the question was a peripheral or a primary participant. They have shown, in other words, that levels of semantic coordination, or the conceptual conventions emergent in interaction are divergent between primary and peripheral participants: an individual indexes these separately to different participants. Insofar as we conceive of context as encoding such emergent ‘semantic ontologies’, this is evidence that an individual keeps track, at least, of this particular aspect of the contexts that she shares with others, separately in the same conversation.

In short, our results indicate that models of dialogue context need to augment individual contributions with information about which participants have processed the contribution and to what degree: they need to index contextual increments on each participant’s information state to specific subsets of the rest of the participants.

8.6 Summary

This thesis is an attempt at articulating a theory of *collective agency* in multi-party conversational dialogue; i.e. how *parties* or alliances, as higher order entities than individuals, can act as unified aggregates in carrying the conversation forward and thus be consequential *units* in how, and on what level, grounding is achieved and how the conversational context develops for each individual.

To this end, the thesis provides a typology of side-participants: primary vs. peripheral, determined by whether the side-participant in question is, at that point, a co-member of the same party as the speaker or the addressee: a consequential distinction which has not been made in the literature. Peripheral participants, on this view, are ‘overhearers’ but independently of ratification or presumed rights of repair/interjection. In this respect, this distinction constitutes a reformulation of the insights offered by Clark and colleagues on how issues of grounding relate to participant role.

Making this distinction has allowed us to provide an account of the idiosyncratic structure of adjacency in multi-party dialogues. That is, it supports the view that adjacency structure is tree-like, and hence can allow for distinct conversational tracks or contexts to be interleaved. Long-distance resolution possibilities for ellipses can be accounted for in this manner. This leads to the replacement of the Conversational Analytic notion of adjacency by the notion of context locality as indexed to specific sub-groups of the participants, based on the distinction between primary

and peripheral participants with systematically distinct possibilities for interpretation. Moreover, we have argued that peripheral and primary participants effect divergent commitments as a result of one and the same utterance. If the intuition is correct that participants actually keep track of these divergent contexts (see below for an experimental design that would test this intuition directly), then we'd be led to the conclusion that DGB-based models of context need to posit multiple DGBs per participant, indexed to subgroups of the rest of the participants present, in order to account for the adjacency structure of conversations of the type we have identified. This does not at all seem to be a trivial task.

People's ability to track conversational contexts over long distances would also have important implications within the psycholinguistic domain. Namely, it would imply that we cannot account for how interpretations are arrived at in egocentric terms alone, as the contexts against which utterances are assessed by an individual cannot be determined without reference to who, or rather what party, has issued the utterance. In other words, utterance interpretation (ellipsis resolution in our case) is often sensitive to who the speaker of that utterance is.

Having said all this, as we've noted, the notion of party-hood is still far from sufficiently explicated. The experimental methods used in this thesis have not allowed for a more local and finer-grained study of why and how alliances take shape or emerge as a product of interaction, e.g. local effects of Split Utterances on the formation of parties (see e.g. Howes, Healey, and Mills (2009) for recent developments along these lines). The parties were, in our experiments, determined in advance and in a sense forced by task-level objectives. Hence, conservatively speaking, the results here only generalise to situations where there's a public and clear agenda of how each topic relates to each participant's concerns and goals (e.g. 2 couples discussing where to meet next with a 5th friend present who is not going) and so they do not generalise fully to e.g. cocktail-party settings, in which alliances can be emergent from the interaction rather than mutually known at the outset. Nonetheless, when they do so emerge, the phenomena identified here are still expected to be at play.

To conclude, what we've dubbed Pragmatic Pluralism (emergence of distinct contexts in the same conversation), could *potentially* have far reaching ramifications for those who try to be *determinate* about the effect of each utterance on each and every participant's mental state. With parties as emergent, and consequential for how each individual's context develops, it might well be that this leads to an inescapable form of semantic non-determinism in dialogue, with uncertainty about uptake of semantic content increasing as the number of participants get larger

thus precluding direct interaction.

8.7 Open Problems and Future Work

8.7.1 Do participants keep track of the distinct dialogue contexts that emerge?

As noted above, our experiments do not show directly that a participant could keep track, separately, of the distinct contexts that we showed she shares with other participants. Below we will sketch the design of an experiment that would provide direct support for this.

Recall our experimental design in chapter 6. Instead of two Directors and one Matcher, in this design we have one Director (D) and two Matchers (M1 and M2), such that the matching of some of the Tangrams is the responsibility of M1 and the rest of M2. The tangrams are marked on the Director's interface as to whom the corresponding Matcher is. The matchers choose from the same set of Tangrams. When a Tangram has been placed by one of the matchers, that Tangram will not subsequently be available to the other.

The Director describes the tangrams to Matchers one by one. Identically to the previous experiment, the fake probe Reprise Fragments (RF) always have distant antecedents, that is, the antecedent always lies in one of the Director's descriptions, within the stretch of talk associated with the previous Tangram. The RFs are in two conditions: same dialogue context and different dialogue context. In the former, the two consecutive Tangrams have been the responsibility of the same Matcher, i.e. D has described both Tangrams to the same Matcher. In the latter condition, the two consecutive Tangrams have been described to different Matchers.

The responses by the single Director to the fake probes would be captured and analysed in the same way as before, for whether the response manifests a distant (correct) or a local (incorrect) interpretation.

The hypothesis would be that the fragments in the different dialogue context condition, get reliably more distant interpretations than those in the same dialogue context condition, i.e. that there would be a systematic conflict between the Director's interpretations of similar RFs, contingent upon whether the RF comes from a locally peripheral or a locally primary participant. This would show that so far as the Director is concerned, while the context that she shared with a primary Matcher during the previous Tangram has 'moved forward', as a result of joint participation with that Matcher during the current Tangram, the context that she shared with a previously primary Matcher, who is locally peripheral, has been 'suspended'. Importantly, this would show that the Director tracks the contexts that she shares with the two Matchers, separately; hence the

need for multiple DGBs.

8.7.2 What is the essence of a party and how/why do they emerge?

We proposed above, that party formations seem to be strongly tied to topic and that they emerge in virtue of relations holding between the participants' publicly known, mutual responsibilities, mutual conversational goals and concerns with respect to that topic. The essence, constitutive of a party, seems to be the sharing of a goal/responsibility among its members, which others in the conversation do not have and this is publicly known. Whether any of this is correct remains an important open question. But it seems that parties, as such, become relevant in the interaction, when the current question under discussion is relevant to that shared goal. But what is the correct notion of relevance here? In other words how do we identify, given a conversational goal, whether an answer to a question under discussion contributes to that goal? If conversational goals themselves are modelled as questions, could this relevance relation be modelled as what Ginzburg (2009) calls 'question dependence or influence'?

Moreover, it is certainly not true that the arrangement in which participants are organised into parties, is always determined at the outset of a topic shift. More often than not parties are emergent from the interaction itself as the various stances that the participants have towards the current topic is gradually revealed. But how so exactly?

8.7.3 Formal treatment

We have presented above, some tentative ideas as to how Ginzburg's KoS could be modified to handle multi-party dialogues of the sorts we have identified in this thesis. The most important change that seems necessary is that conversational update rules should be relativised to participant role - as determined by party membership.

As noted, it seems that questions on QUD, need to have a complex structure, possibly including information about which participants belong to which parties, in addition to who the direct addressees and side participants are (to manage turn-taking obligations). This problem remains unaddressed. Moreover, we have provided very little detail as to exactly how e.g. conversational update rules are to be relativised to participants in different roles, and hence the dynamics of the divergent DGBs remains, for the most part, underspecified. The resulting grounding, querying and assertion protocols, therefore, also remain unanalysed. We have not provided a treatment of how/whether the different forms of intra-party 'conferring' (intra-party elliptical turn assign-

ments, clarification interaction and discussion), in case of grounding failure by one member of a party, lead to backtracking and indeed how backtracking could be accommodated in the model.

Moreover, as we have argued, it is quite likely that multiple DGBs per participant will be needed, in order to account for the pragmatic structuring of multi-party dialogues. If this turns out to be necessary, the dynamics of the multiple DGBs and the update rules responsible for their transitions remain greatly underspecified, though what we have discussed above, does seem to narrow down the choices to be made.

All in all, there is a great deal of work, both empirical and formal to be done.

References

- Allwood, J. (1995). An activity based approach to pragmatics. *Gothenburg Papers In Theoretical Linguistics*, 76.
- Bales, R. F., Strodtbeck, F. L., Mills, T. M., & Roseborough, M. E. (1951). Channels of communication in small groups. *American Sociological Review*, 16, 461-468.
- Bell, A. (1984). Language style as audience design. *Language in Society*, 13, 145-204.
- Bernieri, F., Reznick, J. S., & Rosenthal, R. (1988). Synchrony, pseudosynchrony, and disynchrony: Measuring the entrainments process in mother-infant interactions. *Journal of personality and social psychology*, 54, 243-253.
- Bernieri, F., & Rosenthal, R. (1991). Interpersonal coordination: Behaviour matching and interactional synchrony. In R. Feldman & B. Rime (Eds.), *Fundamentals of nonverbal behavior: studies in emotion and social interaction* (p. 401-432). New York: Cambridge University Press.
- Blom, J. P., & Gumperz, J. (1972). Social meaning in linguistic structure: Code switching in norway. In W. Li (Ed.), *The bilingualism reader*.
- Boden, D. (1994). *The business of talk: Organizations in action*. Oxford, England: Blackwell.
- Branigan, H. P. (2006). Perspectives on multi-party dialogue. *Special issue of Research on Language and Computation*, 4(2-3), 153-177.
- Branigan, H. P., Pickering, M. J., & McLean, J. F. (2005). Syntactic alignment and participant role in dialogue. *Cognition*, 104, 163-197.
- Brannigan, H., Pickering, M., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75(B), 13-25.
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of experimental psychology: Learning, Memory and cognition*, 22, 1482-1493.
- Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, 44, 274-296.
- Cann, R., Kempson, R., & Gregoromichelaki, E. (2009). *Semantics: An introduction to meaning in language*. Cambridge University Press.

- Cann, R., Kempson, R., & Marten, L. (2005). *The dynamics of language: An introduction*. Elsevier Academic Press.
- Carletta, J. C. (1996). Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, 22, 249-254.
- Carletta, J. C. (2006). Announcing the ami meeting corpus. *The ELRA Newsletter*, 11, 2-5.
- Carletta, J. C., Anderson, A. H., & 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.
- Carletta, J. C., Garrod, S., & Fraser-Krauss, H. (1998). Placement of authority and communication pattern in workplace groups: The consequences for innovation. *Small Group Research*, 29(5), 531-559.
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of personality and social psychology*, 76, 893-910.
- Chomsky, N. (1968). *Language and mind*. Harcourt, Brace and World, New York.
- Clark, H. H. (1996). *Using language*. Cambridge University Press.
- Clark, H. H., & Brennan, S. A. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (p. 127-149). Washington: APA Books.
- Clark, H. H., & Carlson, T. B. (1982). Hearers and speech acts. *Language*, 58, 332-373.
- Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84, 73-111.
- Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. Joshi, B. Webber, & I. Sag (Eds.), *Elements of discourse understanding* (p. 10-63). Cambridge: Cambridge University Press.
- Clark, H. H., & Schaefer, E. A. (1989). Contributing to discourse. *Cognitive Science*, 13, 259-294.
- Clark, H. H., & Schaefer, E. F. (1992). Dealing with overhearers. In H. H. Clark (Ed.), *Arenas of language use* (p. 247-274). Chicago: University of Chicago Press.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1-39.
- Colman, M., Eshghi, A., & Healey, P. (2008). Quantifying ellipsis in dialogue: An index of mutual understanding. In *Proceedings of sigdial 2008*. Association for Computational

Linguistics.

- Condon, W. S., & Ogston, W. D. (1966). Sound film analysis of normal and pathological behavior patterns. *Journal of Nervous and Mental Disease*, *143*, 338-347.
- Delamotte-Legrand, R. (1995). Polylogue enfantin ou comment dire a' neuf. In D. Veronique & R. Vion (Eds.), *Mode' les de l'interaction verbale* (p. 397-406). Presses de l'Universite de Provence.
- Drew, P. (1992). Contested evidence in courtroom cross-examination: The case of a trial for rape. In P. Drew & J. Heritage (Eds.), (p. 470-520).
- Duncan, S. D. (1972). Some signals and rules for taking turns in conversations. *Journal of Personality and Social Psychology*, *23*(2), 283-293.
- Eco, U. (1986). Eine palette von grautonen. *Die Zeit*.
- Edelsky, C. (1981). Who's got the floor? *Language and Society*, *10*, 383-421.
- Egbert, M. M. (1997). Some interactional achievements of other-initiated repair in multi-person conversation. *Journal of Pragmatics*, *27*, 611-634.
- Eshghi, A., & Healey, P. G. T. (2007). Collective states of understanding. In B. H. Keizer S. & T. Paek (Eds.), *Proceedings of the 8th sigdial workshop on discourse and dialogue* (p. 2-9). Association for Computational Linguistics.
- Eshghi, A., & Healey, P. G. T. (2009). What is conversation? distinguishing dialogue contexts. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual cognitive science conference* (p. 1240-1246).
- Fay, N., Garrod, S., & Carletta, J. C. (2000). Group discussion as serial monologue or interactive dialogue: The influence of group size. *Psychological Science*, *11*(6), 487-492.
- Fernandez, R., & Ginzburg, J. (2002). Non-sentential utterances: A corpus study. *Traitement automatique des langues*, *43*(2), 13-42.
- Fernandez, R., Ginzburg, J., & Lappin, S. (2007). Classifying non-sentential utterances in dialogue: a machine-learning approach. *Computational Linguistics*, *23*, 397-427.
- Ford, C. E., Fox, B. A., & Thompson, S. A. (2002). Social interaction and grammar. In M. Tomasello (Ed.), *The new psychology of language* (Vol. 2, p. 119-143). Mahwah, NJ: Lawrence Erlbaum.
- Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing 'the' as 'thee' to signal problems in speaking. *Cognition*, *62*, 151-167.
- Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Ef-

- fects of speakers' assumptions about what others know. *Journal of personality and social psychology*, 62, 378-391.
- Garrod, S., & Anderson, A. H. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27, 181-218.
- Garrod, S., & Clark, A. (1993). The development of dialogue co-ordination skills in schoolchildren. *Language and Cognitive Processes*, 8, 101-126.
- Gatewood, J. B., & Rosenwein, R. (1981). Interactional synchrony: Genuin or spurious? a critique of recent research. *Journal of Nonverbal Behaviour*, 6, 12-29.
- Gergen, K. J. (1972). Multiple identity: The healthy, happy human being wears many masks. *Psychology Today*, 5, 31-35.
- Giles, H., & Coupland, N. (1991). *Language: Contexts and consequences*. Pacific Grove, CA: Brooks/Cole.
- Giles, H., & Smith, P. M. (1979). Accommodation theory: Optimal levels of convergence. *Language and social psychology*, 10.
- Ginzburg, J. (1996). Dynamics and the semantics of dialogue. In J. Seligman (Ed.), *Language, logic and computation*.
- Ginzburg, J. (2009). *The interactive stance*. CSLI: Center For The Study of Language and Information.
- Ginzburg, J., & Fernandez, R. (2005). Scaling up from dialogue to multilogue: some principles and benchmarks. In *Proceedings of the 43rd annual meeting of the association for computational linguistics*. Association for Computational Linguistics.
- Goffman, E. (1963). *Behaviour in public places: Notes on the social organisation of gatherings*. The Free Press.
- Goffman, E. (1981). *Forms of talk*. University of Pennsylvania Press.
- Goldinger, S. A. (1998). Echoes of echoes? an episodic theory of lexical access. *Psychological Review*, 105(2), 251-279.
- Goodwin, C. (1979). The interactive construction of a sentence in natural conversation. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (p. 97-121). New York: Irvington.
- Goodwin, C. (1980). Restarts, pauses and the achievement of a state of mutual gaze. *Sociological Inquiry*, 50(2-3), 272-302.
- Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*.

New York, Academic Press.

- Goodwin, C. (1984). Notes on story structure and the organisation of participation. In *Structures of social action* (p. 225-246). Cambridge: Cambridge University Press.
- Grice, H. P. (1957). Meaning. *The Philosophical Review*, 66(3), 377-388.
- Grice, H. P. (1975). Logic and conversation. In *Syntax and semantics: Speech acts* (p. 41-58). Academic, New York.
- Hannah, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: evidence from eye movements. *Cognitive Science*, 108, 831-836.
- Hare, A. P. (1952). A study of interaction and consensus in different sized groups. *American Sociological Review*, 6, 262-267.
- Hare, A. P. (1981). Group size. *American behavioural scientist*, 24(5), 695-?
- Healey, P. G. T., & Battersby, S. (2009). The interactional geometry of a three-way conversation. In *Proceedings of the 31st annual conference of the cognitive science society*.
- Healey, P. G. T., & Mills, G. J. (2006). Participation, precedence and co-ordination in dialogue. In *Proceedings of the annual cognitive science conference*.
- Healey, P. G. T., Purver, M., King, J., Ginzburg, J., & Mills, G. J. (2003). Experimenting with clarification in dialogue. In *Proceedings of the 25th annual conference of the cognitive science society* (p. 539-544). Mahwah, N.J.: LEA.
- Healey, P. G. T., White, G., Eshghi, A., Reeves, A., & Light, A. (2007). Communication spaces. *Computer Supported Co-operative Work*, 17(2-3).
- Heritage, J. (1985). Analyzing news interviews: aspects of the production of talk for an over-hearing audience. In *Handbook of discourse analysis* (p. 95-117). London: Academic Press.
- Homans, G. C. (1950). *The human group*. Thomson Learning.
- Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, 96, 127-142.
- Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59, 91-117.
- Howell, D. C. (1997). *Statistical methods for psychology*. Duxbury Press.
- Howes, C., Healey, P., & Mills, G. (2009, September). A: An experimental investigation into... b: ...split utterances. In *Proceedings of the sigdial 2009 conference* (pp. 79-86). London, UK: Association for Computational Linguistics. Available from

<http://www.aclweb.org/anthology/W/W09/W09-2511>

- Hymes, D. (1972). Models of the interaction of language and social life. In J. Gumperz & D. Hymes (Eds.), *Directions in sociolinguistics*.
- Isaacs, E. A., & Clark, H. H. (1987). References in conversations between experts and novices. *Journal of experimental psychology: General*, 116, 26-37.
- Jefferson, G. (1983). On exposed and embedded correction in conversation. *Studium Linguistik*, 14, 58-68.
- Kalma, A. (1992). Gazing in triads a powerful signal in floor apportionment. *British Journal of Social Psychology*, 31, 21-39.
- Kamp, J. A. W. (1981). A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, & M. Stokhof (Eds.), *Formal methods in the study of language, part 1* (p. 277-322). Mathematisch Centrum.
- Kendon, A. (1970). Movement coordination in social interactions: some examples described. *Acta Psychologica*, 32(2), 100-125.
- Kendon, A. (1976). The f-formation system: Spatial-orientational relations in face to face interaction. *Man environment systems*, 6, 291-296.
- Kendon, A. (1990). *Conducting interaction: Patterns of behaviour in focused encounters*. Cambridge: Cambridge University Press.
- Kerbrat-Orecchioni, C. (2004). Introducing polylogue. *Journal of Pragmatics*, 36(1), 1-24.
- Keysar, B. (2007). Communication and miscommunication: The role of egocentric processes. *Intercultural Pragmatics* 4, 4, 71-84.
- Kraus, R. M. (1987). The role of the listener: Addressee influences on message formulation. *Journal of Language and Social Psychology*, 6, 81-98.
- Krippendorff, K. (1980). *Content analysis: an introduction to its methodology*. Beverly Hills: Sage Publications.
- Kronlid, F. (2008). *Steps towards multi-party dialogue management*. Unpublished doctoral dissertation, Gothenburg University.
- Lakin, J. L., & Chartrand, T. L. (2005). Exclusion and nonconscious behavioural mimicry. In *The social outcast: Ostracism, social exclusion, rejection, and bullying* (p. 279-295). New York: Psychology Press.
- Langenhove, L. van, & Harre, R. (1999). Introducing positioning theory. In R. Harre & L. van Langenhove (Eds.), *Positioning theory* (p. 14-31). Blackwell Publishers.

- Larsson, S. (2002). *Issue based dialogue management*. Unpublished doctoral dissertation, Gothenburg University.
- Leonardi, P., & Maurizio, V. (1990). *Conversazione e terapia. l'intervista circolare*. Raffaello Contina, Milano.
- Lerner, G. H. (1991). On syntax of sentences in progress. *Language in Society*, 20, 441-458.
- Lerner, G. H. (1993). Collectivities in action: Establishing the relevance of conjoined participation in conversation. *Text*, 13, 213-245.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press.
- Levinson, S. C. (1988). Putting linguistics on a proper footing: Explorations in goffman's concepts of participation. In P. Drew (Ed.), *Exploring the interaction order*. University of Pennsylvania Press.
- Lewis, D. K. (1979). Score keeping in a language game. *Journal of philosophical logic*, 8, 339-359.
- Marslen-Wilson, W., & Tyler, L. K. (1976). Memory and levels of processing in a psycholinguistic context. *Journal of Experimental Psychology: Human Learning and Memory*(2), 112-119.
- Matheson, C., Poesio, M., & Traum, D. (2000). Modelling grounding and discourse obligations using update rules. In *Proceedings of the 1st annual meeting of the north american association for computational linguistics*.
- McDowall, J. J. (1978a). Interactional synchrony: A reappraisal. *Journal of Personality and Social Psychology*, 36, 963-975.
- McDowall, J. J. (1978b). Microanalysis of filmed movement: the reliability of boundary detection by observers. *Journal of Non-verbal Behaviour*, 3(2), 77-88.
- Mills, G. J., & Healey, P. G. T. (2006). Clarifying spatial descriptions: Local and global effects on semantic co-ordination. In *Proceedings of the 10th workshop on the semantics and pragmatics of dialogue*.
- Muller, F. (1997). Review of paolo leonardi and maurizio viaro, 1990: *Conversazione et terapia. l'intervista circolare*. *Journal of pragmatics*, 28, 383-412.
- Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal Of Language And Social Psychology*, 21, 337-360.
- O'Connell, D. C., Kowal, S., & Kaltenbacher, E. (1990). Turn-taking: A critical analysis of the research tradition. *Journal of psycholinguistic research*, 19(6), 345-373.

- Parker, K. C. H. (1988). Speaking turns in small group interaction: A context-sensitive event sequence model. *Journal of Personality and Social Psychology*, 54(6), 965-971.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioural and Brain Sciences*, 27, 169-225.
- Power, R. J. D., & Martello, F. D. (1986). Some criticisms of sacks, schegloff and jefferson on turn taking. *Semiotica*, 58, 29-40.
- Purver, M. (2004). *The theory and use of clarification requests in dialogue*. Unpublished doctoral dissertation, Department of Computer Science, King's College London.
- Purver, M., Howes, C., Healey, P. G. T., & Gregoromichelaki, E. (2009, to appear). Split utterance in dialogue: A corpus study. In *Proceedings of the sigdial conference 2009*.
- Reisman, K. (1977). Contrapuntal conversations in an antiguan village. In R. Bauman & J. Scherzer (Eds.), *Explorations in the ethnography of speaking*. Cambridge University Press.
- Sacks, H. (1973). On some puns: with some intimations. In *Sociolinguistics: Current trends and prospects* (p. 135-144). Georgetown University Press.
- Sacks, H., Jefferson, G., & Schegloff, E. A. (1995). *Lectures on conversation*. Blackwell Publishing.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organisation of turn-taking for conversation. *Language*, 50, 696-735.
- Schegloff, E. A. (1972a). Notes on a conversational practice: Formulating place. In D. Sudnow (Ed.), *Studies in social interaction* (p. 75-119). Free Press, Glencoe, IL.
- Schegloff, E. A. (1972b). Sequencing in conversational openings. In J. J. Gumperz & D. H. Hymes (Eds.), *Directions in sociolinguistics* (pp. 346-380). New York: Holt, Rinehart and Winston.
- Schegloff, E. A. (1977). Preference for self correction in the organization of repair in conversation. *Language*, 53(2), 361-82.
- Schegloff, E. A. (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* (p. 71-93). Georgetown University Press.
- Schegloff, E. A. (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. A. (1995). Parties talking together: two ways in which numbers are significant for

- talk-in-interaction. In P. Ten Have & G. Psathas (Eds.), *Situated order: Studies in social organization and embodied activities* (p. 31-42). University Press of America.
- Schegloff, E. A. (1997). Third turn repair. In G. R. Guy, C. Feagin, D. Schiffrin, & J. Baugh (Eds.), *Towards a social science of language* (p. 31-40). Benjamin, Amsterdam.
- Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica*, 4, 289-327.
- Schlangen, D., & Lascarides, A. (2003). The interpretation of non-sentential utterances in dialogue. In *Proceedings of the 4th sigdial workshop on discourse and dialogue*.
- Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In *Handbook of discourse processes* (p. 123-164). Hillsdale, NJ: Lawrence Erlbaum.
- Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, 21, 211-232.
- Searle, J. R. (1969). *Speech acts*. Cambridge University Press, Cambridge.
- Spelke, E. S., Hirst, W., & Neisser, U. (1976). Skills of divided attention. *Cognition*, 4, 215-230.
- Stalnaker, R. (1978). Assertion. *Syntax and Semantics*, 9, 78-95.
- Stasser, G., & Taylor, L. A. (1991). Speaking turns in face-to-face discussion. *Journal of Personality and Social Psychology*, 60(5), 675-684.
- Tan, S., & Moghaddam, F. M. (1999).
 In R. Harre & L. van Langenhove (Eds.), *Positioning theory* (p. 178-194). Blackwell Publishers.
- Traum, D. (1994). *A computational theory of grounding in conversation*. Unpublished doctoral dissertation, Department of Computer Science, University of Rochester.
- Traum, D., & Hinkelman, E. A. (1992). Conversational acts in task-oriented dialogue. *Computational Intelligence, Special Issue: Computational Approaches to Non-Literal Language*, 8(3), 575-599.
- Whitaker, H. (1982). Automaticity. In *Paper presented at the conference on formulaicity*.
- Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, 31, 183-194.