The Function of Intonation in Task-Oriented Dialogue

Jacqueline Claire Kowtko

Thesis submitted for the degree of Doctor of Philosophy
University of Edinburgh
1996

Declaration

I have composed this thesis. The work in it is my own unless explicitly stated otherwise.

Jacqueline C. Kowtko

Acknowledgements

I would like to thank ...

... My supervisors, Steve Isard and Bob Ladd. Steve contributed the original idea for the research and provided unrelenting support (both intellectual and practical) and patience. Bob complemented Steve and provided valuable input as a critic.

... The third member of my 'troika' Ph.D. committee, Ellen Bard, and my various colleagues for their helpful discussions and encouragement: particularly Simon Garrod, Gwyneth Doherty-Sneddon, Anne Anderson, Jean Carletta, Cathy Sotillo, Bethan Davies, and other members of the Human Communication Research Centres in Edinburgh and Glasgow.

... My family and friends who were interested in my work and gave encouragement.

... The UK government for an Overseas Research Student Award, the Human Communication Research Centre in Edinburgh (Director, Keith Stenning and former Acting Director Ellen Bard) and the Centre for Speech Technology Research (Director, Steve Isard) for continuity of support.

The writing of this thesis ended with a measure of joy I had not anticipated. (Psalm 23)

Considering the many different English accents I might have studied, it was a great pleasure to do research on the Glasgow accent. Listening to dialogues from the Map Task Corpus was often akin to enjoying a baroque concert. Glaswegian is *music* to my ears.

To Jenne and Michael

Abstract

This thesis addresses the question of how intonation functions in conversation. It examines the intonation and discourse function of single-word utterances in spontaneous and read-aloud task-oriented dialogue (HCRC Map Task Corpus containing Scottish English; see Anderson *et al.*, 1991). To avoid some of the pitfalls of previous studies in which such comparisons of intonation and discourse structure tend to lack balance and focus more heavily on one analysis at the expense of the other, it employs independently developed analyses. They are the Conversational Games Analysis (as introduced in Kowtko, Isard and Doherty, 1992) and a simple target level representation of intonation. Correlations between categories of intonation and of discourse function in spontaneous dialogue suggest that intonation reflects the function of an utterance. Contrary to what one might expect from reading the literature, these categories are in some cases categories of exclusion rather than inclusion.

Similar patterns result from the study of read-aloud dialogue. Discourse function and intonation categories show a measure of correlation. One difference that does appear between patterns across speech modes is that in many instances of discourse function, intonation categories shift toward tunes ending low in the speaker's pitch range (e.g. a falling tune) for the read-aloud version. This result is in accord with other contemporary studies (e.g. Blaauw, 1995). The difference between spontaneous and read results suggests that read-aloud dialogue – even that based on scripts which include hesitations and false starts – is not a substitute for eliciting the same intonation strategies that are found in spontaneous dialogue.

Contents

Chapter 1

Introduction

When people converse, they can utter the same words in different ways and make them mean something completely different. For example, a passerby might approach me on the street and ask directions. In the course of our interaction, I might at some point say “Right” meaning a direction, at another point say “Right” meaning that I have understood what that person is saying, and at yet another point say “Right” inquiring if that person has understood my directions. The passerby easily distinguishes the meaning, or utterance function, carried in what I say.

Various studies show that intonation helps listeners make a host of different distinctions. It helps us to identify cue words (e.g. Hirschberg and Litman, 1991), discourse boundaries (e.g. Swerts and Geluykens, 1994), the ends of turns (e.g. Cutler and Pearson, 1986), and other phenomena in conversation. When the passerby hears my utterance “Right”, he or she extracts some meaning from the intonation. What is not clear is the exact nature of the connection between intonation and function.

The subject of intonation function has been a concern to different areas of research for different reasons. Much of the early work on this subject was carried out to instruct teachers of the English language to foreigners (e.g. Palmer, 1922). It has largely taken an approach which pays close attention to intonational form and provides rather impressionistic views of meaning in terms of speaker attitude, emotion, and intention.

In more recent decades, since technological advances enabled an automatic means of obtaining fundamental frequency (F0, which can be used to check auditory judgements and other aspects of intonation analyses) various researchers have made efforts to take a more empirical approach to both sides of the problem. Studies of particular discourse structures or boundaries (e.g. Hockey, 1992; Hirschberg and Litman, 1991) look carefully at specific discourse contexts and try to determine how intonation categories correlate. Other studies (e.g. Pierrehumbert and Hirschberg, 1990) look carefully at intonational form but try to be more systematic about interpretations of meaning than the earlier impressionistic work.

This thesis addresses the link between discourse function and intonation contour. It also considers the difference between intonation strategies in spontaneous and read-aloud speech.

Such a study of intonation function has implications in the field of computer dialogue systems.[1] The desire to understand intonation (and other aspects of prosody) is increasing as researchers have begun to recognise a potential need for computer systems to encode prosodic features. A better knowledge of intonation may be able to help a dialogue system in different ways. For instance, it may help to syntactically disambiguate phonetically identical phrases (e.g. Price *et*

[1] Computer dialogue systems, or speech understanding systems, integrate natural language processing, speech recognition, and a component of dialogue control. They accept speech as input toward accomplishing a specified task, and provide speech or written language as output (e.g. the dialogue control system in Young and Proctor, 1991). The ultimate goal for some dialogue system builders seems to be creating a self-contained, mechanised conversationalist that would converse with humans with such skill that it would be difficult to distinguish one from a human. Such a machine, or the hopes for it, resemble fragments of that classic monster created by Frankenstein (Shelley, 1818), able to converse with its creator and think as a human. Frankenstein's creation became too powerful for the good of creator and world, surpassing its original, intended abilities. I would hope that designers of speech systems will not try to create artificial life, but will instead enjoy studying one of the most wonderful capacities given to the human race, language.

al., 1991). It may also help to parse phrases by perceiving phrase boundaries (e.g. Pagel *et al.*, 1995; Veilleux *et al.*, 1990) and help to constrain grammar selection by identifying utterance function types (Taylor *et al.*, 1996).

Although speech recognisers have achieved relatively high levels of success within limited domains without encoding prosodic information,[2] as efforts are being re-directed towards dialogue systems, issues in prosody are eliciting more interest. Since the systems perform a role similar to a human participant in dialogue, their designers see various issues in dialogue as important to understand. These issues include intonation and dialogue structure.

The answer to whether intonation strategies are the same in spontaneous and read-aloud dialogue may have implications for researchers using read-aloud speech to train and test speech recognisers. Such systems need to be trained on prosodic coding, and it might be helpful to know whether they could use read speech to control quality for intonation training materials. Phoneme recognisers train on read speech because its content and quality can be controlled more easily than in spontaneous speech. Few systems train on spontaneous speech (the recogniser described in Taylor *et al.*, 1996, fine tunes phone models from spontaneous speech). Intonational differences between read and spontaneous speech are potentially interesting for recognition systems that utilize this suprasegmental information to identify certain linguistic structures. For this reason the study in Chapter 7 involves read-aloud dialogue which arises from people reading detailed transcripts of conversations they participated in previously. The transcripts are very detailed and include disfluencies such as false starts, repetitions, and hesitations. Other studies examining read versus spontaneous data have 'cleaned up' the transcripts for the reading task (e.g. Ayers, 1994). While this may seem a good idea, it eliminates some of the charac-

[2] e.g. The SPHINX speaker-independent recognition system (Lee, 1990) encodes no suprasegmental information (phoneme and word duration are encapsulated in segments) yet it achieves 93.7% word accuracy in continuous speech taken from the DARPA/TIMIT database (Price *et al.*, 1988) with a vocabulary of 997 words and a word-pair grammar of complexity 60 (one of approximately 60 words may follow an identified word).

teristics of spontaneous speech and makes the read dialogue more like written language in that it has no disfluencies.

1.1 Approach of the Thesis

In approaching a study of intonation function, one must be careful to maintain balance and use as comprehensive a framework for discourse structure as for intonation. In order to avoid the imbalance of one analysis taking precedence over another, the present work treats intonation and discourse analyses as equally important. In the examination of intonational function, a robust, independently motivated method of functionally analysing discourse is needed. The system which fulfills this need is one developed by Kowtko, Isard and Doherty-Sneddon (1992; described in Chapter 4).[3] The studies in this thesis examine single-word utterances in dialogue. This limits the complications which longer utterances pose. For example, it avoids problems such as identifying the location of pitch accents and intonational phrase boundaries. Nevertheless, a comprehensive framework is required. Very short utterances present problems to many contemporary analyses of intonation because those analyses were designed to apply to longer phrases. Chapter 5 introduces a new system.

The main hypothesis which is tested in the intonation study is based upon an assumption shared by others who work on intonation in dialogue (e.g. Hockey, 1991, 1992; Litman and Hirschberg, 1990; McLemore, 1991), namely that discourse function correlates significantly with the choice of intonation contour.

[3] Since subsets of the present work first appeared in conference papers (e.g. Kowtko, 1992), other work based on our discourse distinctions (as in Kowtko *et al.*, 1992, and Carletta *et al.*, forthcoming) has begun to do in other languages what this project started for (Glasgow) Scottish English. See Grice *et al.* (1995).

1.2 Goals of the Thesis

This thesis has two basic goals. First, it seeks to understand how intonation functions in dialogue by merging a functional analysis of dialogue with an independent analysis of intonation contour. It reduces the domain of study to single-word utterances in task-oriented dialogue (from the HCRC Map Task Corpus which contains Scottish English, Anderson *et al.*, 1991). This study is in Chapter 6. Second, it compares intonation strategies in two types of dialogue: spontaneous and read-aloud. This study is in Chapter 7.

The studies find correlations between discourse and intonation categories in spontaneous and read dialogue. Although the patterns are similar across dialogue modes, the intonation in read-aloud dialogue includes more tunes which fall and end low in pitch (similar results from a study involving a subset of the data are reported in Kowtko, 1995). Read dialogue differs slightly in a way which suggests that speakers use different strategies in the two modes. Results suggest that read dialogue may not be suitable for training recognition systems unless the developers are willing to accept a margin of error.

1.3 Organisation of the Thesis

Because the studies in Chapters 6 and 7 cover two currently separate disciplines, discourse analysis and intonational phonology, background is provided for each area. Chapter 2 describes some of the approaches to analysing conversation which provide a setting for the *Conversational Games* Analysis (from Kowtko *et al.*, 1992) which appears in Chapter 4. Chapter 4 also describes the Map Task Corpus and presents experiments between coders using the Games Analysis which establish acceptable levels of coder agreement and thus validate the useability of the analysis. Chapter 3 reviews approaches to intonational phonology. It also discusses previous work on the function of intonation, introduces western Scottish intonation (the Glasgow accent), and discusses the differences between spontaneous and read speech, as a precursor to Chapter 5 which introduces a new analysis of intonation.

The two strands meet in Chapters 6 and 7 which describe studies of intonation function in single word utterances taken from spontaneous and read-aloud dialogue, respectively. Chapter 7 also includes a comparison of individual data points across dialogue modes. This establishes that categories of discourse function in the Games Analysis are as good at predicting intonation correlations as the speakers themselves are, supporting the validity of the Games Analysis as a representation of discourse function. Conclusions are presented in Chapter 8.

Appendices A and B provide an example of a coded dialogue and excerpts from the raw results, respectively.

Chapter 2

Previous Work on Discourse

As background to the discourse framework of the intonation study in Chapter 6 (i.e. the Conversational Games introduced in Chapter 4), this chapter reviews the literature which motivates the development of a new analysis. Several areas of discourse research address issues relating to the organisation of conversation. Speech Act Theory (Austin, 1962; Searle, 1969) lays the philosophical foundations. Other theories develop their approaches from viewpoints of different academic fields. For example, Conversational Analysis lies within sociology, Discourse Analysis within linguistics, Collaborative Theory within psychology, and goal-directed approaches within computing disciplines. The analyses of conversation take different styles of approach, from identifying general principles of interaction to creating taxonomies of structure.

2.1 Some Definitions

The literature varies somewhat in how it uses terms relating to discourse. To avoid confusion, this thesis will maintain the following definitions.

Discourse refers to speech or written language involving more than one utterance or sentence. (This is consistent with the dictionary definition of *discourse* as a speech or lecture in either written or spoken form.) The term encompasses language as uttered spontaneously and language that has been crafted, e.g. by a speech writer or playwright. It also includes sentences which linguists make up to illustrate points in a theory.

Conversation refers to discussion between two or more persons.

Dialogue refers to discussion between two persons.

Discourse analysis refers generally to the field crossing linguistics and related disciplines which examines function and form in discourse. Although this label is often applied to work on monologue,[1] in the present thesis it will focus on work on dialogue. There is a specific sense of the term which defines its own area of research, but this will be introduced later in a new section (2.4).

2.2 Speech Act Theory

Speech Act Theory (Austin, 1962; Searle, 1969) provides a vocabulary and foundation for analysis of utterance function. Austin and Searle make several contributions which are pertinent to the analysis of spoken dialogue: analyses of speaker intention, meaning, and the beliefs of the speaker and hearer. Searle says simply (p. 20), "We need to distinguish what a speaker means from certain kinds of effects he intends to produce in his hearers."

The basic premise of Speech Act Theory is that speech is action. By uttering something we are doing something. Austin distinguishes three types of acts which can simultaneously be performed in issuing an utterance. He states (1962: 109),

> We first distinguished a group of things we do in saying something, which together we summed up by saying we perform a *locutionary act*, which is roughly equivalent to uttering a certain sentence with a certain sense and reference, which again is roughly equivalent to 'meaning' in the traditional sense. Second, we said that we also

[1] e.g. Rhetorical Structure Theory by Thompson and Mann, 1987.

> perform *illocutionary acts* such as informing, ordering, warning, undertaking, &c., i.e. utterances which have a certain (conventional) force. Thirdly, we may also perform *perlocutionary acts*: what we bring about or achieve *by* saying something, such as convincing, persuading, deterring, and even, say, surprising or misleading. Here we have three, if not more, different senses or dimensions of the 'use of a sentence' or of 'the use of language' (and, of course, there are others also).

The acts differ in terms of which aspect of speech is being addressed: meaning in locution, intention in illocution, and effect in perlocution. As an example of the three acts, Austin refers to the utterance "Shoot her!" (pp. 101-102):

Locution He said to me 'Shoot her!' meaning by 'shoot' shoot and referring by 'her' to *her*.

Illocution He urged (or advised, ordered, &c.) me to shoot her.

Perlocution He got me to (or made me, &c.) shoot her.

The acts also differ in terms of the directness of the connection to the utterance itself. The locutionary act is most directly related to the speech uttered since it arises from components of the actual utterance, and the perlocution is the least direct, being the (often non-linguistic) result of the utterance.

Austin considers not only the effect of an utterance but its context. "The particular persons and circumstances in a given case must be appropriate for the invocation of the particular procedure invoked" (p.15). However, his context is related to action (situational), and he neglects to address the issue of how linguistic context affects the interpretation of utterances.

The illocutionary act carries across a *force*, i.e. the function. Austin proposes classes of illocutionary force: verdictives (e.g. judgement, appraisal), exercitives (e.g. voting, warning), commissives (e.g. promising, committing), behabitives (e.g. apologising, challenging), and expositives (e.g. replying, assuming). The issue of illocutionary force is something which Searle (1969) carries further. Searle (pp. 24-25) adds a dimension of analysis to Austin's by distinguishing between action (utterance) and meaning (proposition):

Utterance acts performed by uttering words (morphemes, sentences),

Propositional acts performed by referring and predicating,

Illocutionary acts performed by stating, questioning, commanding, promising, etc., and

Perlocutionary acts as in Austin (1962), relating to the consequences or effects upon the hearer.

Searle presents an analysis of eight frequently-used types of illocutionary act: *request, assert, question, thank, advise, warn, greet,* and *congratulate.* He discusses the types of rules which apply to the use of these acts. One of these rules is a preparatory condition for the act. The preparatory condition for the act *thank (for)* is that an act benefits the speaker and the speaker believes that the act benefits the speaker. The analysis for *thank* further specifies that the speaker's intention must be to convey to the hearer the gratitude felt toward the hearer. The speaker must believe that a benefit has been received and want to express appreciation for that benefit.

It is primarily the illocutionary act, as defined by Searle, which finds its way into theories of discourse analysis and dialogue production. The illocutionary act addresses the effects upon the hearer (e.g. in terms of belief)—an important consideration in the study of interactive speech.

Austin addresses the issue of hearer interpretation (which has implications for the acceptance phase of interactive discourse; see Collaborative Theory in Section 2.5) in his reference to *uptake.* Austin considers an illocutionary act complete only when it is interpreted with the speaker's intention, otherwise the act is invalid (p.116-117): "Unless a certain effect is achieved, the illocutionary act will not have been happily, successfully performed. ... So the performance of an illocutionary act involves the securing of *uptake*". Searle does not make this specification.

Compare Edmondson (1981) who proposes a 'hearer-knows-best' principle, that it is the hearer's interpretation of the force of an utterance rather than the speaker's intended force which matters. If the hearer is incorrect in the interpretation, the speaker (who is by now the hearer) will self-correct.

2.3 Conversational Analysis

Conversational Analysis refers generally to the field of sociolinguistic research which takes the viewpoint of the ethnographer. It developed from ethnomethodology and focuses on conversational interaction. The ethnography of speaking comprises an organizational study of the norms of interaction – expected behaviours or rules to which people conform. This approach tends to focus on specific process within interactive discourse, e.g. the precise timing of closing sequences (Jefferson, 1973), rather than overall structures of entire conversations. Sacks, Schegloff, and Jefferson are the pioneers in the field of Conversational Analysis, producing work which elucidates some of the internal structures of conversation.

Sacks, Schegloff and Jefferson (1974) propose the existence of a turn-taking mechanism. The issue is one of sequencing in conversation. They propose rules of turn-taking and repair mechanisms for turn-taking violations and errors. Their rules such as 'current speaker selects next', 'one speaker at a time', and 'first starter has rights' help account for the orderly exchanges in conversation.

Schegloff and Sacks (1973) introduce the idea of adjacency pairs. An adjacency pair consists of two adjacent utterances spoken by different persons. There are many different pair types, e.g. question-answer, greeting-greeting, offer-acceptance/refusal. Schegloff and Sacks propose the following rule: when speakers recognise the first part of a pair they should try to produce the second part from a pair type in which the first is a member. The pairs are essentially comprised of initiation and response.

The adjacency pair is foundational in the ethnomethodological model. It is the structural organization to which speakers orient in their turn-taking, and it shapes their expectations. If the second part of a pair is missing, speakers look for alternate interpretations (cf. Grice's, 1975, co-operative principle, Section 2.7). The adjacency pair, however, does not account for second parts which occur not in adjacent position but perhaps one turn away.

2.4 Discourse Analysis

The term *Discourse Analysis*, in addition to its general sense, refers to a specific area of research. This area involves work on speaker interaction but differs from Conversational Analysis in several ways (see Coulthard, 1985). One salient difference between the two is that Discourse Analysis attempts the explanation of a broad, overall theory of interactive discourse, while the other concentrates on specific processes within interactive discourse.

In the 1970's, a group of researchers (e.g. Sinclair, Coulthard, Brazil and Johns) at Birmingham University carried out several large research projects on various aspects of interactive discourse. (The group is known as the "Birmingham school".) Their various projects attempt to gain an overall understanding of interactional function. In particular, Sinclair and Coulthard's (1975) detailed study of classroom discourse contributes a framework of conversation in terms of units at different levels.

Sinclair and Coulthard approach the description of discourse from a structural linguistic point of view. As a linguistic analysis of grammar identifies units and their sequence, they present an analogous description of discourse. Their system has four criteria:

1. A finite descriptive apparatus (categories)
2. Symbols or terms which map clearly to data elements
3. A comprehensive descriptive system – to cover all data
4. At least one impossible combination of symbols (limited syntax)

They identify five ranks on the discourse level (as opposed to the grammar level), from top to bottom: lesson, transaction (roughly change of topic/focus), exchange (e.g. adjacency pair), move, and act. Their ranking is a structure of linguistic action which reflects the sequential character of the interaction between teacher and pupils. They identify two types of exchange, boundary and teaching exchanges. Moves can be opening, answering, follow-up, framing, or focusing types. Particular exchanges consist of particular moves. For instance, a teaching exchange has an Initiation (an Opening move), an optional Response (an Answering move) and optional Feedback (a Follow-up move). The lowest rank is the act. Three major acts, elicitation, directive, and informative, appear in classroom discourse. There is also a non-verbal act. The act "silent stress" is "realized by a pause, of the duration of one or more beats, following a marker" (p. 43).

Sinclair and Coulthard propose an analogy between grammatical and discourse structures: the move is to the sentence as the act is to the clause. Therefore, they consider the move to be the basic unit.

Sinclair and Coulthard find that teaching exchanges have three slots – Initiation, Response, and Feedback (E=IRF). Teachers and pupils alike understand this sequence as a part of the discourse structure in a classroom.

Stubbs (1983) extends the analogy of discourse to grammar and makes a competence versus performance distinction. He argues for the concept of well-formedness in discourse – that a category may be appropriate or inappropriate and speakers recognise it. For example, speakers know that I is followed by R and then F. Speakers might agree or disagree as to well-formedness.

Sinclair and Coulthard acknowledge that grammatical form and discourse function have a relationship which may change depending upon the situation. The example they provide is the interrogative "What are you laughing at?" which often surfaces as a command but sometimes a question. They propose a few rules to help identify an interrogative as a command versus a question.

Critics of the approach that Sinclair and Coulthard take (called the "rules and units" approach by Taylor and Cameron, 1987) point to the limitations of rule-governed analysis and classification of interaction into categories. Taylor and Cameron believe that the rules and units framework blinds us in our grasp of the conceptual, theoretical, and methodological realities in conversation. (This is understandable because in the framework we do not try to understand everything, merely the regularities.) They discuss the principle of intersubjectivity, the idea that participants in conversation work toward a goal of mutual awareness and understanding, i.e. similar models of the conversation. They propose abandoning the principle because it ignores the reality that individuals bring their own viewpoints and maintain their own perceptions of a conversation and its progress.[2] This question of differing viewpoints and interpretations is addressed at least in part by Collaborative Theory, specifically content grounding. Participants, although they are assumed to want to build similar models of conversation, must work on agreeing that they understand each other and thus bring their viewpoints into alignment.

2.5 Collaborative Theory

Collaborative Theory (Wilkes-Gibbs, 1993), as the name suggests, focuses on the collaborative aspects of conversation. "This theory treats speaking and listening as fundamentally collaborative processes, and conversing in any language as a strategic, collective activity." (p.1 of ms.). While, in one sense, the other approaches to discourse analysis involve collaboration, e.g. the second part of an adjacency pair should be appropriate to the first, they do not make explicit the nature of collaboration.

Collaborative Theory originates in the work on grounding presented by Clark and Schaefer (1987; see also Clark, 1996: chapter 8). People collaborate while conversing to build mutual belief – common ground. *Contributions* form the

[2] In this respect they agree with Kreckel (1981). Kreckel suggests that interpretation of speech acts is different for individuals because each has a different point of view – different experience of linguistic conventions (one aspect of cultural norms) in which that interpretation is fostered.

units of conversation. Each contribution consists of *presentation* and *acceptance* phases which occur in the effort to establish the mutual belief that everyone has understood the content of the contribution. In the presentation phase, the speaker presents an utterance whose content he or she wishes to convey. In the acceptance phase, the hearer communicates that it is understood, grounding the content (with perhaps an explicit verbal signal or merely continued attention). Repair occurs in the acceptance phase until both speakers are happy with the resultant contribution.

Other collaborative processes noted by Wilkes-Gibbs include continuation (cf. the FILL feature in the Conversational Games Analysis, Section 4.3.4) and completion of another's utterance. People also collaborate with respect to referring expressions. In this, they may accommodate one another. Experts accommodate novices by using better descriptive expressions (in a photo-ordering task, Wilkes-Gibbs, 1993).

2.6 Communication Accommodation Theory

Communication Accommodation Theory (CAT), and the more specific Speech Accommodation Theory (SAT) are notions based on the observation that speakers often change their verbal behaviour to accommodate their speaking partner (Giles *et al.*, 1991). Speakers often exhibit linguistic and interactional accommodation. For example, a speaker might adapt speech rate or backchannel[3] frequency to match the partner's. This basic strategy which people exhibit in communication is called convergence. White (1989) describes one example of convergence. She finds that American speakers converge with respect to frequency of backchannel responses when speaking to Japanese, whose natural frequency is quite high.[4]

[3] Backchannel is grounding-related feedback such as "uh-huh".

[4] In 30-minute conversations, Americans increase from 28 backchannels per conversation with other Americans to 46 backchannels when speaking to Japanese. Japanese, on the other hand, only reduce slightly when talking to Americans: 60 backchannels compared with 68 when speaking to other Japanese.

Accommodation relates to a person's need for social approval. A speaker may use dis-associating techniques to maintain a distance from the partner (e.g. to mark different social class). Features which have been found to converge through accommodation include utterance length, speech rate, information density, vocal intensity, pausing frequency and lengths, response latency, gesture, and posture.

2.7 Pragmatics

Pragmatics considers meaning in terms of function within a context. The most relevant contribution of pragmatics to the present discussion of discourse analysis is from Grice (1975). Grice proposes the Co-operative Principle: conversations are co-operative events. Participants adhere to this principle as they verbally interact. Several maxims jointly support the principle. They are characterised in terms of Quantity, Quality, Relation, and Manner.

According to the Co-operative Principle, a speaker may flout a maxim, i.e. appear to violate it, and the hearer as a result will to try to reinterpret the utterance in such a way that it would not violate the maxim. One such example involves telling a joke. The speaker may state something blatantly false, and thus appear to violate the maxim of quality. The hearer seeks an alternate interpretation – 'joke' instead of 'literal truth' – because the participants are co-operating (cf. Stubb's 1983 assessment of well-formedness, in Section 2.4).

2.8 Computational Approaches to Dialogue Analysis

Three computational approaches to dialogue analysis are discussed in this section. The first was developed more or less concurrently with the present research and addresses some basic issues concerning interaction. The second and third were developed previously and serve as foundations for the Conversational

Games Analysis (introduced in Chapter 4).

2.8.1 Conversational Acts

Traum and Hinkelman (1992), present a theory of Conversation Acts at four levels based on an analysis of task-oriented dialogue: argumentation acts (similar to exchanges), core speech acts, grounding acts (e.g. initiations, acceptances, repairs), and turn-taking acts. They discuss the need for a new unit of discourse, the *discourse unit*. It is an exchange composed of grounding acts, in which content grounding occurs (as in Clark and Schaefer, 1987). In the discourse unit, a traditional speech act forms the initiation (the presentation phase), and another grounding act is expected to follow, such as an acceptance which indicates that the speech act has been understood (the acceptance phase). An acceptance may consist of an explicit acknowledgement such as “okay” or a content-bearing or meaningful response which indicates that the initiation is accepted. One example of an acceptance is a reply to a question. Non-verbal acceptances such as nods and eye gaze, are not addressed, presumably because the dialogues on which they base their analysis did not involve eye contact. (The participants were in different rooms and the verbal channel was the sole means of communication between participants; see Allen and Schubert, 1991, p.29).

An utterance may be coded as more than one act. For example, a reply to a question may function as an acceptance (a grounding act) and as an initiation (another grounding act). That is, the same utterance which implicitly grounds the other speaker’s question also begs its own grounding acceptance.

Sometimes utterances appear to lack grounding. This leaves a discourse unit incomplete. Traum and Hinkelman (p.13) deal with the problem of a non-grounded utterance by suggesting that it is automatically grounded. Because the utterance is not challenged and because the content appears to be incorporated into later plans in the dialogue, by default it is accepted.[5]

[5] This is reminiscent of work by Cohen and Levesque. Cohen and Levesque (1987) argue

Traum and Hinkelman do not address the frequency of grounding (or backchannel) utterances. Some studies in psychology and sociology address backchannel frequency, and indicate that it is linked to social issues.[6] A theory of dialogue should be able to handle the presence and lack of backchannel. Traum and Hinkelman leave an ungrounded utterance as an incomplete discourse unit. This creates an anomaly within their structure because the lack of explicit grounding does not always affect the progression of the conversation, yet the incompleteness of the discourse unit indicates that it might.

Like most of the above authors of discourse analyses, Traum and Hinkelman ignore the issue of useability.[7] Another weakness of their analysis is that turn-taking acts are rather poorly defined. They allude to intonational and other prosodic cues but avoid talking about specifics and do not refer to any of the literature (some of which is in the next chapter, Section 3.4).

2.8.2 Conversational Procedures

Power (1974, 1979) examines how conversation arises from underlying non-linguistic goals. He describes a computer programme which generates conversation between two simulated robots performing a simple task. The task involves sliding bolts on doors, opening them, and moving through them. In order to achieve goals in this limited world, one robot has to elicit help from the other robot. Power proposes *Conversational Procedures* to elicit such help. For example, if one robot wants to move through a door, he needs first to find out

with respect to uptake (Austin, 1962) that explicit recognition of an utterance's force is unnecessary because the force of an illocutionary act can be judged from speaker intentions, based on mutual beliefs. They base force recognition not on speech actions but basic principles of action (goals) and intention (belief). Effects of an act are analysed in terms of inferences and rational thought.

[6] E.g. Duncan and Fiske, 1977, find that backchannel rates such as nods are slightly higher for female than male subjects. This topic is covered within Accommodation Theory (e.g. Giles, Coupland, and Coupland, 1991; see Section 2.6 above).

[7] Sinclair and Coulthard (1975: 61) allude to the problem of replicability in their provision of annotated transcripts for aspiring analysts.

whether the door is bolted on the other side, where the other can see it. The first conversational procedure would be to ANNOUNCE his intention to move. The he would secure the other robot's cooperation through an AGREEGOAL procedure.

The procedures assign roles to each robot and instruct each regarding how to interpret and produce specific utterances, so that at any point both know when the interaction has successfully reached completion. Consider the procedure ACHIEVEGOAL which occurs after the robots agree to cooperate on a particular plan to accomplish something such as sliding the bolt on a door. Subtasks in ACHIEVEGOAL include performing a direct act (e.g. sliding the bolt), testing whether a state of affairs has been achieved, and selecting a plan.

Power allows procedures to nest, as robots sometimes find it necessary to initiate a new procedure (e.g. ask a question) in order to proceed through the current one.

Power's programme advances upon previous dialogue generation systems in that it represents goal-related motivation or intention behind an utterance.

2.8.3 Interaction Frames

Houghton (1986; also Houghton and Isard, 1987) adopts Power's notion of Conversational Procedure, simplifying the procedures into four *Interaction Frames* for his own robot agents who live in an artificial world similar to that which Power describes. Speakers have goals which they want to accomplish, and they elicit the help of a partner through conversation. Goals surface in the model as plans for exchanges of information, 'dialogue games' (Houghton and Isard, 1987). Given the communicative intention to do something related to the goal, one can plan the utterance within the current context.

Houghton claims that four interaction frames, GET-ATTENTION, MAKE-KNOWN, FIND-OUT and GET-DONE, are sufficient to accomplish simple goals. After the initiation of a frame, the responder must recognise the type of interaction started and generate an appropriate response.

Interaction frames specify the following:

- The type of goal the interaction is to be used for (e.g. having someone tell you information),
- Precondition tests for attempting the interaction (e.g. that the addressee likely knows the information),
- Procedures for the non-verbal activities that the interaction may require for either participant (e.g. searching memory), and
- The type of reply expected of the addresssee.

Conversational plans help the participants choose which interaction frame to access. Participants look to see if any frames are open (which require closing) before choosing a new one to open.

The descriptions of conversational interaction developed by Houghton and Power are intended for computer-generated dialogue. Chapter 4 introduces the Games Analysis (Kowtko *et al.*, 1992) which has adapted and revised Houghton's Interaction Frames, tailoring them to actual human dialogue. The analysis is used to comprehensively represent dialogue arising from a map task for the studies in Chapters 6 and 7.

Chapter 3

Previous Work on Intonation

This chapter introduces terminology and reviews work on intonational phonology, in preparation for the intonation analysis of single-word utterances in Chapter 5. It also discusses the accent in (Glasgow) Scottish English. Finally, it reviews some of the links between intonation and discourse structure and some of the differences between spontaneous and read speech.

3.1 Introduction to Intonation

Intonation refers to the tonal or pitch variation which occurs in speech. It is suprasegmental, and as such is superimposed on a series of vowels and consonants (or strings of syllables).

An *intonational phonology* is an abstract representation of intonation. It distinguishes intonational categories in such a way as to identify the minimal units. Some approaches consider the units to embody meaning (e.g. Pike, 1945; Bolinger, 1985; Gussenhoven, 1984; Ladd, 1996; and the standard British tradition which associate units with attitudes and emotions) and others describe simply categorical units with no intended sense of meaning (e.g. Pierrehumbert, 1980; 't Hart *et al.*, 1990). In most accounts of English intonation, intonationally prominent syllables play an important role.

Prominence in a syllable or word refers to the salience of that item compared to its surrounding context. Prominence can be characterised in different ways.

Traditionally, three features are identified which contribute to its perception. One feature is pitch (fundamental frequency or F0). Scholars from as early as the 17th century identify this feature. Butler (1634, p.54) introduces pitch in a discussion of *accent* (meaning general prominence) from a book on English grammar:

> Tone is the natural and ordinary tune or tenor of the voice: which is to rise or fall, as the Primary points shall require: and therefore it dominateth the voice, High or Low.[1]

He also identifies loudness (amplitude or intensity), a second feature of prominence. A third feature of prominence is length (duration). These three features lend to the perception of prominence in a syllable.

Bolinger (1958) distinguishes *pitch accent* from *stress*. The former relates to prominence and the latter, word stress. His experiments with pitch, duration, intensity, and their position within a word cause him to conclude that "the primary cue of what is usually termed STRESS in the utterance is pitch prominence. [...] Intensity is found to be negligible both as a determinate and as a qualitative factor in stress" (p.149). However, the acoustic distinction is not that clear cut.

The traditional view that stress does not involve pitch while intonation does is not entirely correct. Beckman (1986) examines stress (or stress accent as she calls it) in different languages and finds that the relationship between stress and intonation is not easy to define. Loudness, or intensity, appears to be a production cue for stress, but "the defining feature is relative loudness as reflected not in the peak amplitude level but rather in varying quantities of total amplitude" (p. 200). That is, stress is related more to the overall energy produced in the syllable.

The relationship between stress and pitch accent underlies the link between metrical phonology and intonational phonology. The field of metrical phonology

[1] Primary points consist of eight punctuation marks which determine tone, sound (loudness) and pause.

addresses issues concerning the description and the prediction of the location of stress in utterances. Liberman and Prince (1977) and Selkirk (1984) present two influential approaches to metrical phonology based on the idea that the meter of a phrase can be represented by a hierarchy (or grid) consisting of binary weak-strong distinctions. Various rules are postulated which assign stresses of different strengths to syllables in phrases. The strongest stress plays an important role in the description of intonation. As pitch accent and word stress are related, intonational phonologists often consider metrical phonology an integral part of understanding the whole of intonation (e.g. Pierrehumbert's, 1980, use of the work by Liberman and Prince). Indeed, Sluijter and van Heuven (1995) argue against completely independent tonal and metrical structure in phonological theory, as the two are integrally linked.

The type of accent of interest to intonational phonology is pitch accent. It will be referred to merely as accent, below.

3.2 Intonational Phonology

There are various approaches to representing intonation in English. The phonological descriptions largely fall into two groups, British and American systems.

3.2.1 British Approaches

"Standard British"

The notion of a tone group as the structural and functional unit of intonation, and the concept of a nuclear tone, forms the basis of the "Standard British" approach. It stems mainly from work aimed at teaching spoken English. Various terms have been used refer to the group of tones in a phrase: intonation group (Armstrong and Ward, 1931; Kingdon, 1958), tone group (Palmer and Blandford, 1935; O'Connor and Arnold, 1973), and tonic unit (Crystal, 1969).

By definition, the tone group consists of at least one accented syllable which carries a tone, the nuclear tone. The tone group for longer utterances consists of a head, nucleus, and tail (Palmer, 1922). The nuclear syllable bears the main prominence and is usually the final accented word in the group. The head comprises all syllables (accented, stressed, or unstressed) prior to the nucleus, and the tail, all syllables (stressed or unstressed) following the nucleus. If the head contains other accented syllables, their prominence is subordinate to the nucleus.

Others within the British tradition subdivide the head into prehead, head, and body (Kingdon, 1958; Schubiger, 1958). The head is redefined as the first accented syllable of the tone group. The prehead consists of everything before the head, and the body subsumes the stretch of syllables between the head and nucleus. The more recent, generally accepted analysis (e.g. O'Connor and Arnold, 1973; Crystal, 1969) divides the tone group into prehead, head, nucleus and tail. Each non-nuclear structure may include a stretch of utterance, (instead of being limited to a single syllable, as with Kingdon's and Schubiger's head). Thus, the prehead extends from the beginning of the tone group to the first accented syllable, the head begins with the first accented syllable and ends on the syllable preceding the nucleus, and the tail follows the nucleus. Combinations of different types of these structures (preheads, heads, nuclei plus tails) determine the classes of tone groups. Some of the observed functions of these types are described in Section 3.4.

There are two types of tone group. One is relatively minor and groups words with close grammatical connection (the boundaries are marked with a "|" by O'Connor and Arnold, 1973). It may or may not end in a pause. The other is relatively major and ends with a definite pause, separating utterances that are not closely connected, e.g. at the boundary of two sentences (marked with a "||").

The tone of a nucleus is characterised as a contour – a change in pitch. O'Connor and Arnold (1973) identify seven nuclear types: Low Fall, High Fall, Rise-Fall, Low Rise, High Rise, Fall-Rise, and Mid Level (neither rise nor fall).

There are problems with the standard British approach. An important one is that it is not always easy to identify the nuclear accent. For example, there may be two equally prominent pitch accents which vie for nuclear strength; or, the boundary of the tone group may be unclear, obscuring the location of the nucleus. The problem of identifying the nucleus also has repercussions in making more difficult the process of correctly identifying and classifying the head and tail.

Modified "Standard" Representation

Knowles *et al.* (1996) solve the problem of the nucleus by eliminating it altogether. They adapt the standard system (as presented by O'Connor and Arnold, 1973) to produce a transcription for a corpus of English speech. Three major changes are incorporated. They

1. discard the notion of a nuclear accent
2. distinguish all accents (except compounds) as having high and low variants
3. define high and low in terms of the immediately preceding pitch level (not pitch range)

Knowles *et al.* identify major and minor tone group boundaries and distinguish nine tones: High Fall, Low Fall, High Rise, Low Rise, High Level, Low Level, High Fall-Rise, Low Fall-Rise, and Rise-Fall.

This representation seems to offer the best of the British tradition, in that it avoids the theoretical complications of the nucleus and provides a workable system to use on large corpora of speech. However, as is argued below in Section 3.2.2, a description of pitch accent in terms of target levels captures the acoustic data better than a description in terms of pitch movement, such as the British tradition offers.

3.2.2 American Approaches

The American descriptions take a different approach to intonation.

Levels

Structural linguists in early to mid-20th century America generally characterise intonation in terms of pitch levels. They analyse the height of accented syllables as they appear on several tiers relative to the speaker's pitch range. For example, Pike's (1945) influential analysis identifies four distinct levels relative to each other: 1 *very high*, 2 *high*, 3 *mid*, and 4 *low*. Like the British analysts of his day, Pike classifies sequences of the tones which represent the intonation of utterances and discusses how intonation carries a speaker's attitude (see Section 3.4 on intonation function). This work sets a foundation for later American phonologies.

Accent Targets

Pierrehumbert's (1980) phonology represents intonation in terms of target levels, not contours.[2] In this respect, it continues the historical American sense of identifying levels of pitch accent. Unlike the old levels systems, however, it uses only two levels, low (e.g. valley) and high (e.g. peak). The levels are relative to each other, not necessarily relative to the pitch range. For example, a high tone in a phrase which is spoken with a low, narrow pitch range could actually be lower than a low tone in a phrase with wide pitch range (cf. Pierrehumbert's, 1980, discussion of a downstepped high being lower than an initial low accent).

Because only the two tones are used, implementation rules which link pitch traces to tone sequences must be carefully constructed. These rules are accent-specific. Different accents have different types of pitch contours, and with two tones, only a limited number of mappings between tone combinations and pitch contours are possible. Pierrehumbert's framework specifies one type of accent (American English), although the components of her two-tone system can apply to other accents and other languages.

[2] Bolinger's (1985) analysis of American intonation represents pitch movement in a manner more similar to the British system, as contours. Like Pierrehumbert's phonology, described below, it lacks adequate representation of level tunes.

Pierrehumbert chooses a theory of levels over one of pitch movement (configurations) because it allows theoretical distinctions not possible in the other approach. For example, a theory of pitch movement considers a terminal fall to a speaker's baseline (lowest pitch) and a fall to mid-level which is then sustained ('vocative contour' or stylised fall; Ladd, 1988) to be two instances of a falling tune. Both tunes start relatively high at about the same F0 level. A target level theory calls the two tunes different entities. In Pierrehumbert's system, H*L-L% (fall to baseline) and H*+L H-L% (downstepped high tone which in the ToBI version of Pierrehumbert's work would be H* !H-L% fall to mid-level) distinguish two contours which a system based on pitch configurations cannot. Ladd (1996) points out additional phonetic evidence for the target levels approach, citing various studies in which productions hit certain F0 targets, e.g. that baseline lows maintain a particular value despite different pitch range.

The analysis identifies two types of phrases, intermediate phrases (Beckman and Pierrehumbert, 1986) and intonational phrases. Their components correspond to different parts of phrases: initial boundary tones correspond to phrase beginnings, pitch accent to accented syllables, phrase accent to intermediate phrase endings, and boundary tones to intonational phrase endings.

Pierrehumbert describes an inventory of seven tunes: H*, L*, L*+H, L+H*, H*+L, H+L*, and H*+H. (The tune H*+L triggers downstep.) Tone on an accented syllable is marked with an asterisk or 'star' (*). There are two phrase accents, H- and L-, two boundary tones, H% and L%, and two initial boundary tones H% and L%. The grammar of a tune has components as shown in Figure 3.1. An intonational phrase optionally begins with an initial boundary tone. Then it iterates through the pitch accents (identified using Liberman and Prince's (1977) theory of metrical phonology), using one of the the seven tunes for each accent. Intermediate phrases end with a phrase accent. Intonational phrases end in a boundary tone.

The phonology is tailored to American English. Pierrehumbert explicitly mentions (p.47) that the sequence L*H-L% never occurs as rise-plateau-falling tail or rise-fall. (See Figure 3.2). As will be discussed below in Section 3.3, the

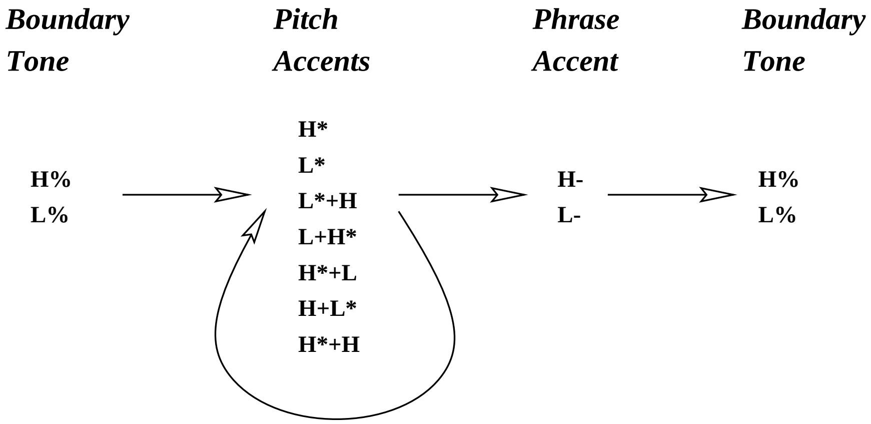

Figure 3.1: Components of Pierrehumbert's (1980) Intonational Phrase Grammar

characteristic Glaswegian rise which is rise-plateau-falling tail therefore cannot be represented in Pierrehumbert's description.

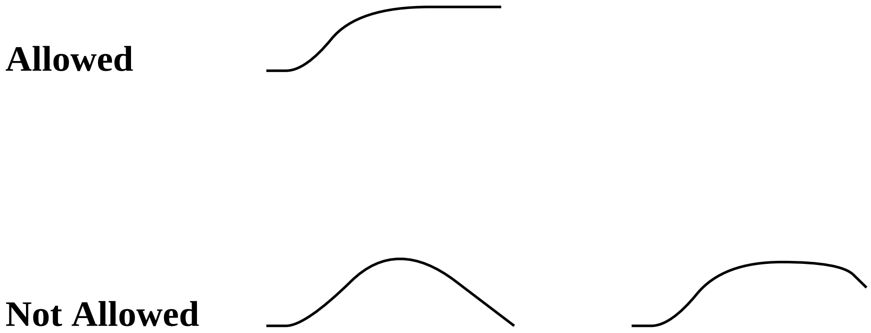

Figure 3.2: The Allowable and Non-Allowable Shapes of L*H-L%

Pierrehumbert's system has difficulty in representing level tunes. It realises H*H-L% as high level while L*H-L% is low followed by higher level plateau. L*L-L% is a gradual fall. There is no corresponding low level.

Pierrehumbert (p.c.) holds that a single-word phrase with level accent can be represented by a sequence of high accent, high phrase and high boundary tone (H*H-H%) for a "high level tone" and low accent, low phrase and low boundary tone (L*L-L%) for a "low level tone", but this presents problems. The same representations (H*H-H% and L*L-L%) in Pierrehumbert's original analysis map to non-level contours. Figure 3.3 illustrates the problem. The sequence L*L-L% can represent a low pitch accent followed by a low phrase tone followed by a low boundary tone, making one low, level contour. This sequence can also represent a low tone followed by dropping phrase and boundary tones. Similarly, the transcription H*H-H% can represent a high, level contour or a high, rising contour. Nothing in the analysis makes it possible to distinguish the two possible interpretations for the sequences L*L-L% and H*H-H%. The utterances used in the studies require distinct representations for rises, levels, and falls.[3]

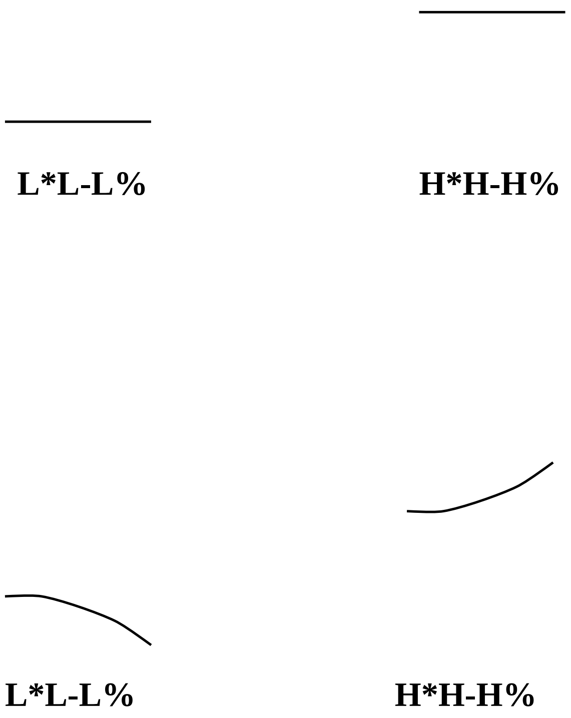

Figure 3.3: Dual Mapping of L*L-L% and H*H-H%

3.2.3 A Proposed Standard: ToBI

Silverman *et al.* (1992; see also Beckman and Ayers, 1994) propose a standard for prosodic marking, Tones and Break Indices (ToBI). It has two main components: a break index tier and an intonation tier. The break index tier marks word boundary strength on a scale of 0 to 4. The scale ranges from phonetic

[3] McLemore (1991) also encounters the problem of levels in Pierrehumbert's system and solves it by modifying the transcription and introducing a critical mark for sustained, level tone value (T:).

linking such as assimilation across boundaries (index 0) to intonational phrase boundaries (index 4). Its tonal component, the intonation tier, is based upon Pierrehumbert's work with a few changes made. ToBI disposes of the L% initial boundary tone and alters the representation of downstepped tones in pitch accents and phrase accent. It removes the tones H*+H and H*+L, which triggers downstep in the following H tone, and adds the diacritic '!' to downstepped H tones (!H*, L+!H*, H+!H*, and !H-).

ToBI exhibits similar problems to Pierrehumbert's work in its representation of level tunes and Glaswegian rises.[4]

3.2.4 Approaches at a More Phonetic Level

The RFC Analysis

Taylor (1992) proposes a more phonetic analysis of intonation contours which can be used for speech synthesis and recognition purposes. He suggests that intonational phrases can be transcribed as units which rise, fall, or connect (RFC). These are the components with which intonation contours can be represented, but the RFC notation makes no reference to presence or lack of pitch accent. RFC refers to an acoustic-phonetic level. An intonational phrase can be represented by numerical parameters which specify RFC type, duration, and amplitude.

Taylor and Black (1994) build on Taylor's work and produce an analysis which is closer to the phonological domain: *tilt*. They identify intonational 'events' which can be computationally represented by four parameters: amplitude, duration, position, and tilt. Although the intonational event is a more phonological unit, distinctions between different events are defined by computational parameters, and the system is best suited to a computational application.

[4] One might circumvent the problem by devising a Glaswegian ToBI similarly to Grice *et al.* (1995) who have applied it to the description of German (See Grice *et al.*, 1996), Italian, and Bulgarian accents. Japanese, Korean, and indeed Glaswegian ToBI are all under development.

The Dutch System

The Dutch analysis developed at IPO originates in a bottom-up, experimental-phonetic approach to describing melody ('t Hart *et al.*, 1990). It represents contours as accented units which have been analysed, resynthesised, and stylised. Perceptual experiments establish their validity as intonation units. The resultant contours are called close-copy stylizations. They decompose pitch movement into perceptual features: Direction (rise, fall), Timing (early, late, very late), Rate of change (fast, slow), and Size (full, half).

A repertoire of ten perceptually relevant pitch movements have been established for Dutch (represented by numerals 1 through 5 and letters A through E). They combine to form the intonation contours found in Dutch. Figure 3.4 shows an example of a contour with three melodic shapes. Similar inventories have been established for other languages, including British English (de Pijper, 1983).

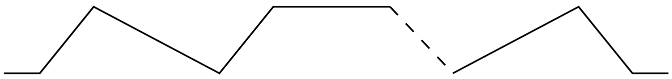

Figure 3.4: The sequence /1D/ /1B/ /4A/ ('t Hart *et al.*, 1990, p.160)

De Pijper (1983) uses IPO analysis-resynthesis techniques in a melodic model of English intonation. He identifies 8 pitch movements for English, represented by the following perceptual features:

- Direction (rise, fall)
- Steepness (steep, gradual)
- Range (full, half)
- Position (early, middle, late)

Position is location of pitch movement with respect to syllable. The values of these features have not only impressionistic reality but also precise phonetic representation. Two parameters, *slope* and *duration* are used. Computational correlates have been established through perceptual experiments. For example, a fast, early rise has an increment rate of 50 semitones/second with a duration of 120 milliseconds.

De Pijper's representation (also the Dutch IPO representation) is generally at an acoustic-phonetic level, although it incorporates some phonological distinctions such as the rise and fall high/low distinction. De Pijper associates his model with Halliday's (1970) phonological description, a more or less standard British analysis.

The Dutch/IPO analysis differs from Taylor's and Black's in a few respects. The Dutch system is accent-specific. It identifies distinct tunes – shapes of intonational components that are significant in a specific accent. Taylor's and Black's events locate the presence of a tune. The latter analysis is more general and universally applicable to different accents.

3.3 The English Accent in Glasgow

A number of British accents, especially urban accents in northern areas, are characterised by non-standard intonation[5]. One of them is Glaswegian. The accent in Glasgow is similar to these other northern urban accents in that it involves a great number of final rising tunes which typically occur in these non-standard accents where RP has a falling tune (Cruttenden, 1986, 1995). While in RP the final accent commonly falls in pitch, and any subsequent unaccented syllables remain low, in Glaswegian it rises in pitch and subsequently maintains a high plateau sometimes with a slightly falling tail. Macaulay (1977) identifies the contour as a rise-fall.

Macaulay reports that the most common tunes in an excerpt of conversation by two Glaswegian speakers, accounting for 77% of the tunes, are as summarised in Table 3.1.

It is quite likely that the first two are actually instances of the same tune. The

[5] The 'standard' accent is Received Pronunciation (RP), the accent on which the standard British tradition is based. It is spoken mostly in southern England.

Table 3.1: Macaulay's (1977) Observed Common Glaswegian Tunes ($N = 263$)

Tune	*Frequency*
(Lower) Mid to Higher Mid	17%
Mid to Higher Mid to Mid	36%
gradually down from (Higher) Mid; some syllables may form a Level succession	24%

presence of additional syllables past the final accented syllable may account for the slight drop in pitch from higher mid to mid level.

The Glaswegian final rise differs from rises in other English accents in that the rise often begins immediately before the accented syllable and ends after it, on subsequent unaccented syllables. In other words, the accented syllable occurs somewhere during the actual rise. This poses a possible theoretical problem for target level analyses such as Pierrehumbert's which associate an accent with a particular target (e.g. L* for a clear valley or turning point in pitch). Ladd (1996) addresses this problem. He suggests that the problem lies with the narrow understanding of "starred tone" in Pierrehumbert's analysis of American English. So perhaps a solution would allow the starred tone to mean something broader for different accents, e.g. that in American English L* pinpoints a visible dip in F0 whereas in Glaswegian it may represent the beginning of a rising tone. If Pierrehumbert's framework is to be used, I prefer an analysis which retains the definition of a starred tone and analyses the Glasgow rise as a starred LH – L*H or (LH)* or some other representation which does not force the accented syllable to associate with either the L or H. The L and H are real in terms of F0, but the actual prominence usually occurs between them. Ladd additionally notes that if the sequence L*H-L% is to be applied to the Glaswegian rise, rules of phonetic realisation of phrase and boundary tone targets will have to be tailored to specific accents, so L*H-L% which is a stylised rise in American

English will appear as a rise-plateau-slight fall in Glaswegian English. He calls this an unsatisfactory resolution, as it gives identical phonological analyses to markedly different contours in two mutually intelligible varieties of the same language. Yet, it appears to be the best solution for adapting Pierrehumbert's intonational phonology of (American) English. Such issues are currently under investigation for the Glasgow accent and may be published at a later date. For purposes of this thesis, the Glaswegian rise which occurs in single-word utterances is represented in terms of accent target levels (see Section 5.3).

Unstressed syllables in Glaswegian often maintain a high pitch. Brown, Currie and Kenworthy (1980) describe Glasgow speech as unlike RP and Edinburgh accents in that it has a raised base-line from which stressed syllables scoop down in pitch, yielding the rising-contour pattern characteristic of west coast Scottish speech. In the other accents the base-line is low, and stressed syllables are perceived as higher in pitch. While this idea of a raised baseline is not actually appropriate or correct in Glaswegian, it helps the reader to understand that Glaswegian fundamentally differs from the other accents.

3.4 Intonation Function in Discourse

It is a widely held belief that intonation has meaning in itself (e.g. the British tradition; Bolinger, 1989; Gussenhoven, 1984; Ladd, 1996). This meaning is related to speaker attitudes and emotions. Scherer *et al.* (1984) find a correlation between intonation categories and attitudes and emotions. Collier (1993) reviews some findings regarding the expressive function of intonation. Ladd (1996: 38-40) advocates the 'Linguist's Theory of Intonational Meaning'. The central idea is that the elements of intonation have meaning. "These meanings are very general, but they are part of a system with a rich interpretive pragmatics which gives rise to very specific and often quite vivid nuances in specific contexts." This theory of meaning underlies much of the work on intonation and discourse structure.

Studies on the link between intonation and discourse structure tend to take one of two approaches. Some studies focus on discourse and look for intonation correlates. Others focus on intonation phenomena and look for discourse correlates. Some of the findings of these approaches, discourse-focused and intonation-focused, are described below. One problem which these approaches share is that they tend to ignore the theoretical framework of the one area (intonation or discourse) on which they do not focus. The description of that area is often general and may be impressionistic.

It is important to realise that specific results of such studies might be valid only for the particular accent examined. In most studies, this accent is RP or American.

3.4.1 Discourse-Focused Approaches

These approaches involve focusing on discourse functions and examining intonation from that point of view. These studies usually have well-defined discourse functions, and they report any interesting F0 effects relating to particular discourse events. A variety of these functions are described immediately below.

Various distinctions in discourse such as given versus new information, contrastivity, and other phenomena are conveyed by intonation features (e.g. rising v. falling tunes; see Cruttenden, 1986; Walker, 1992). Discourse boundaries are also indicated by intonation features (Hirschberg and Pierrehumbert, 1986; Swerts and Geluykens, 1994). When a speaker begins a new topic in conversation, or when a news reporter makes a direct quote, he or she uses an expanded or higher pitch range (Brown, Currie and Kenworthy, 1980; Ayers, 1994; Grosz and Hirschberg, 1992).

Speakers mark information units by manipulating the relative height of pitch peaks (e.g. Swerts and Geluykens, 1994; Nakajima and Allen, 1993). They also mark information flow by manipulating the distribution of pauses and their relative length (Swerts and Geluykens, 1994; Grosz and Hirschberg, 1992). Coherence between succesive utterances (in Swedish) can be described in terms of downdrift of F0 peaks and valleys across the whole tone group (Bruce, 1982).

Macafee (1983) observes that a continuing part of the discourse is often signalled by rising pitch movement from a non-low position, in Glaswegian. (This, however, is not the Glaswegian 'continuation' marker; see Section 8.2.)

Intonation and other aspects of prosody help to signal turn-taking. Brown *et al.* (1980) report that continuation of a turn may be marked with non-low terminal accent (in Edinburgh English). Pauses and repetition also indicate the end of a turn. Speakers (of RP) may signal the end of a turn by lowering pitch and loudness (Beattie, Cutler and Pearson, 1982) and using a rapid downstepped contour at the end (Cutler and Pearson, 1986). Expanded pitch range cues a new turn and a correction (Ayers, 1994).

French and Local (1986) find that pitch height plus loudness associate with turn-competitive interruptions, as opposed to those interruptions that are not competing for control.

Particular speech acts may involve substantial use of one particular tune. For example, telephone greetings often have a high-mid tone (Liberman and McLemore, 1992).

Phrasing and accent can be used to disambiguate the function of syntactically ambiguous words, e.g. cue and non-cue words (in monologue and dialogue) more readily than textual analysis. For example cue use of the word "now" is characterised by the accent occurring alone in a phrase, or in initial position with L* accent or deaccented (Hirschberg and Litman, 1987, also 1991; Litman and Hirschberg, 1990).

Sag and Liberman (1975) suggest a method using intonation to disambiguate indirect speech acts from direct ones. They identify sentence tunes which help to separate questions from suggestions and other indirect speech acts.

Utterance types associate with different intonation patterns. McClure (1980) finds that in western Scottish intonation the typical pattern for wh-questions (e.g. "Why will he be arriving on Monday?") is similar to that in simple statements (e.g. "He'll be arriving on Monday."): The single phrase peak descends to the last stressed syllable where a sudden rise occurs, followed by a sudden drop. Wh-questions differ from statements in two respects:

1. In a wh-question, the pitch ends lower than the level at the beginning of the utterance, unlike statements which end at the same level as they begin.

2. The coincidence of pitch peaks with stressed syllables is less exact in question than statement patterns.

McClure characterises the distinctions between statements, wh-questions and yes-no questions in terms of the intonation curve at or near the first pitch accent and the last pitch accent:

Type	*First Accent*	*Last Accent*
Statement	high fall	high fall
Wh-Questions	high fall	mid fall
Yes-No Questions	mid fall	high fall

McClure finds that true rising intonations in phrase-final position (as one finds in RP) are very rare in the Glasgow accent. When they do occur, they tend to associate with high levels of emotion.

3.4.2 Intonation-Focused Approaches

Approaches which focus on intonation tend to have a fairly coherent theory of intonation but lack any sense of a theory of discourse structure. They often consider one intonation category at a time and search for any salient features of discourse which correlate. This section briefly describes a number of these studies.

As the earlier analyses of intonation originate in texts for teaching spoken English (even Butler, 1634, provides grammar instruction), the authors attempt to associate particular meanings with different intonation patterns to help readers understand when the patterns should be used. They tend to describe meaning in terms of attitude, emotion, and other effects and draw sometimes impressionistic conclusions. For example, O'Connor and Arnold (1973) identify some of the functions of various tunes. The ‘high drop’ tone group (low pre-head + high head + high fall) indicates completeness and definiteness. The height of the fall adds power (e.g. to an imperative, “Will you do that.”) In a question tag, the high drop demands agreement. In a negative statement it may indicate a skeptical attitude. It may indicate mild surprise. Kingdon (1958) gives impressionistic correlates of intonational meaning, e.g. he characterises an utterance as being a mocking or impatient statement or being an insistent question. Palmer (1922) indicates that intonation helps to classify methods of expression, i.e. to separate types of speech acts such as greeting, reassuring someone, asking permission to do something, interrupting and protesting.

Pike (1945) and Bolinger (1985, 1989) also talk about meaning in terms of emotion and intention. Bolinger (1989) discusses some of the correlations between particular sentence types and intonation patterns, while repeatedly emphasizing that “no intonation is an infallible clue to any sentence type” (p.98). He observes, for instance, that the B+AC (high rise plus quick drop and then low rise) contour is often used in echo questions.

Gussenhoven (1984) takes a different perspective with regard to meaning and considers the status of information in terms of its status in the background, or shared understanding, of a conversation. He links a falling tune to the process of adding a variable to the background (new information). A rise says nothing definite about whether the entity is in the background or not. A fall-rise selects a variable from background (given information).

Cruttenden (1986) seeks universal meaning for intonation types. He suggests that rising tunes have “open” meanings (i.e. generally non-assertive and continuative) and falling tunes have “closed” meanings (i.e. generally assertive and non-continuative).

More recently, Pierrehumbert and Hirschberg (1990) discuss the function of tunes from Pierrehumbert’s phonology of American English. For instance, they claim that the L* accent marks items that the speaker intends to be salient but not to form part of what the speaker is predicating in the utterance. Hobbs (1990) goes one step futher and describes the abstract character of what intonational elements signify. He concludes that H*/L* signals new or not new, a shift from L or H to H* or L* indicates a kind of correction or accommodation to what the hearer might have believed the status to be, and a H suffix indicates that the status is still an open question.

Brazil (1975) introduces the concept of key in conversation. Key is decided by the relative height in pitch of the first prominent syllable of the tonic segment. It corresponds to certain general meanings: high key is contrastive; mid key is additive or neutral; low key is equative. The key of one speaker's termination (final tone unit) and the initial key of the respondent tend to exhibit concord.

McLemore (1991) examines three types of phrase-final accents and the discourse functions which they accomplish. In general, rising tunes connect, level tunes continue, and falling tunes segment. This, is, in a sense, little different from what early 20th century English grammar books for foreigners do in their impressionistic descriptions of intonation function. Although McLemore works systematically using recorded conversations and monologues and provides specific instances of the types of contexts which qualify as connecting, continuing, and segmenting, there is no discourse framework in which her analysis is couched. Descriptions of function depend upon the context, which is left up to the reader to assess. For example, among her conclusions are that phrase final levels marks continuation or boredom, continuity without participation or interruption by the audience, provides background, creates suspense, is used for recurrent business, etc. Phrase-final rise (indicating connection or non-finality) can manifest in turn-holding, phrase subordination, or intersentential cohesion.

Hockey (1991) admits to settling upon an arbitrary system of discourse classification in an analysis of dialogue arising from a task in which one participant tells the other how to reproduce a design made with coloured paperclips. Although her results concerning the cue phrase *okay* correlate with McLemore's (89% of *rising* contours occur where the speaker was *passing* up a turn and letting the other person continue; 86% of *level* contours serve to *continue* an instruction; 88% of *falling* contours mark the *end* of a subtask), the discourse categories used are not corroborated by an independent judge and the results are thus not replicable.

McLemore's and Hockey's work lack overall discourse theories. The studies in Chapters 6 and 7 combine independently motivated intonation and discourse analyses in a study of how intonation correlates with discourse structure.

3.5 Differences between Read and Spontaneous Speech

A variety of prosodic phenomena have been examined and compared in spontaneous and read speech (monologue and dialogue). This section summarises some findings.

Segment duration may cue speech mode. Spontaneous speech appears to be faster than read speech (in terms of number of syllables per second, Blaauw, 1995) because it deletes phonetic segments. In fact the two types of speech cover similar number of segments per second (Bernstein and Baldwin, 1985). Perception of the identity of utterances as spontaneous or read speech is helped by articulatory duration, segmental duration, and in particular liquid ([**r**], [**l**]) duration (Remez *et al.*, 1991).

Factors which help distinguish spontaneous and read speech include boundary type (minor, major, none), tune type (falling, rising), pre-boundary lengthening, and pause (Blaauw, 1994). Pauses for breath occur at 100% of the grammatical junctures (e.g. ends of syntactic phrases) in read speech and 69% of cases in spontaneous speech (Goldman-Eisler, 1968). Blaauw (1995) finds that virtually all major boundaries are marked intonationally in both speech modes. Only half of the minor boundaries are marked in both modes. Phrase-internal boundaries are common in spontaneous speech but not read speech. In spontaneous speech they usually precede highly informative words.

Pausing frequency is sensitive to the requirements of the verbal task. It increases with semantic complexity and decreases with the learning or rehearsal of speech (Goldman-Eisler, 1961).

Related to pausing is the structure of intonational phrases. Spontaneous speech has a greater number of shorter phrases than the equivalent text in read speech. It has more phrases, fewer words per phrase, and fewer accents per phrase than read speech (Ayers, 1994).

In both spontaneous and read speech (dialogue), expanded pitch range accompanies the beginning of new topics (Ayers, 1994).

Pitch range differs in the two modes. Some studies find that overall, pitch range is bigger in read speech (Ayers, 1994; Blaauw, 1995; Johns-Lewis, 1986). Johns-Lewis also finds that mean F0 is higher in read speech. Remez *et al.* (1985), Hieronymus and Williams (1991), and Blaauw (1995) find evidence which suggests the contrary, that pitch range is bigger in spontaneous speech. This difference is apparently not rooted in a monologue versus dialogue distinction. Ayers and Johns-Lewis use speech from spontaneous conversation and readings of (edited) scripts of that conversation or acting (dialogue with a fictitious partner). Blaauw, Remez *et al.* and Hieronymus and Williams use spontaneous monologue and read sentences excised from it. The difference may lie in factors related to the purpose or the setting of the speech. Blaauw finds that pitch range is smaller in spontaneous speech from interviews (perhaps more similar to conversation) and larger in spontaneous speech from instruction monologues. Remez *et al.* and Hieronymus and Williams collect their spontaneous speech by asking subjects various open-ended questions, some of which elicit instruction from the subject.

The repertoire of tunes appears to be similar in both speech modes, yet there are some differences. Bruce and Touati (1992) observe (impressionistically) that the same inventory of pitch patterns appear in Swedish read laboratory speech and spontaneous dialogue. They describe an example of these pitch patterns, downstepping and non-downstepping contours as they relate to instances of focal accent. Some of the differences between tunes across speech modes involve falling tunes. Spontaneous speech exhibits steeper falls than read speech (Hieronymus and Williams, 1991). Blaauw (1994) notes that many more falling boundary tones appear in read speech and more rising boundary tones in spontaneous speech. Levels are largely shared.

The study in Chapter 7 seeks to establish the similarity between intonation (pitch accent) strategies in spontaneous and read dialogue.

Chapter 4

Analysis of Spoken Discourse: Conversational Games

This chapter presents the *Conversational Games* Analysis (Kowtko, Isard & Doherty-Sneddon, 1992; Carletta *et al.*, 1995), a theory which represents discourse structure at two levels. It was developed with the intent to analyse utterance function in dialogue and to be used in various academic studies. In this thesis (Chapters 6 and 7) it is applied to the study of how intonation functions in dialogue.

4.1 Types of Spoken Discourse

Ideally a method of discourse analysis will be able to cope with different types of spoken discourse, from linguistic exchanges between two individuals to discussions between several participants. As a constraint on the domain, the Conversational Games Analysis in its development was applied to task-oriented dialogue. In particular, dialogues from the HCRC Map Task Corpus (described below in Section 4.2) were used. There are several reasons for this choice. Firstly, the corpus is large (128 dialogues), readily available, and has transcripts and audio recordings of high quality. Secondly, the corpus was designed to be the object of several types of cognitive and linguistic studies, and the discourse analysis was intended to be used by researchers other than the developers. Thirdly, task-oriented dialogue is found in human-computer interactive speech understanding systems. It was hoped that this analysis would be useful in improving such systems. It is reasonable to assume that a computer speech understanding system developed on dialogue oriented around one particular task would more easily be adaptable to handle dialogue arising from another task than a system developed on another type of discourse, e.g. discursive conversation.

Although the Games Analysis was developed on task-oriented dialogue, it is possible to apply the theory to conversations between three or more persons. In this chapter the theory will be presented with respect to dialogue.

For research purposes task-oriented dialogue is preferred over other types of less constrained dialogue for a number of reasons. On a practical note, task-oriented dialogue is easier to analyse. Vocabulary is limited. Speaker roles are more structured. Turns between speakers tend to be more regular. Units of conversation are easier to recognise, as they relate to goals within the task. Knowing the speakers' goals allows the analyst to assess more easily the intent and function of the speaker's contribution. Analysing intention sometimes resembles reading a speaker's mind. Task-oriented dialogue lends itself to an analysis of intention since the task is known and goals can be deduced.

4.2 Map Task Corpus

The dialogues involved in the development of the Conversational Games Analysis are part of the Human Communication Research Centre (HCRC) Map Task Corpus (See Anderson *et al.*, 1991). The Corpus is a collection of 128 dialogues centred around a task involving a map game and is available on audio tape (or digital audio tape – DAT) and written transcript. An example of a Corpus dialogue can be found in Appendix A. Section 4.2.10 contains an excerpt.

4.2.1 Purpose

The Corpus project arose from a recognised need for some method of eliciting speech which would allow multi-faceted analysis of conversation and human interaction. The goal was to produce a body of data useful for a variety of academic studies. Researchers from the fields of linguists, psychology, and artificial intelligence were among those involved in the design and creation of the Corpus. The dialogues needed to be carefully constrained by design yet spontaneous in nature. They also needed to be recorded with high quality.

4.2.2 The Task

The task around which the dialogues centre is a simple map game (designed by Brown *et al.*, 1984). Two participants play. Each participant has a map with various landmarks on it and cannot see the other person's map. One map has a path. The participants are told that different explorers have drawn the maps. The person with the map that has a path (the instruction giver) tells the other person (the instruction follower), how to reproduce the path on the other map. There is no time restriction. The task ends when the instruction follower has drawn a path from start to finish.

4.2.3 Materials

A set of 16 pairs of maps was used in the Corpus (See Figure 4.1 for a sample pair of maps.). The maps themselves have intricate design. The maps within a pair are similar but differ slightly in terms of the match between landmarks. A landmark consists of an iconic illustration with a name written below it. All pairs of maps include several landmarks which match in name and icon. They also include at least one of each of the following types of differences: landmarks which are slightly mismatched in name (e.g. dutch elm v. dead tree), a landmark which is present on one map and absent on the other, a landmark which appears twice on the giver's map and only once on the follower's map, and different landmarks with contrasting names (e.g. crane bay v. green

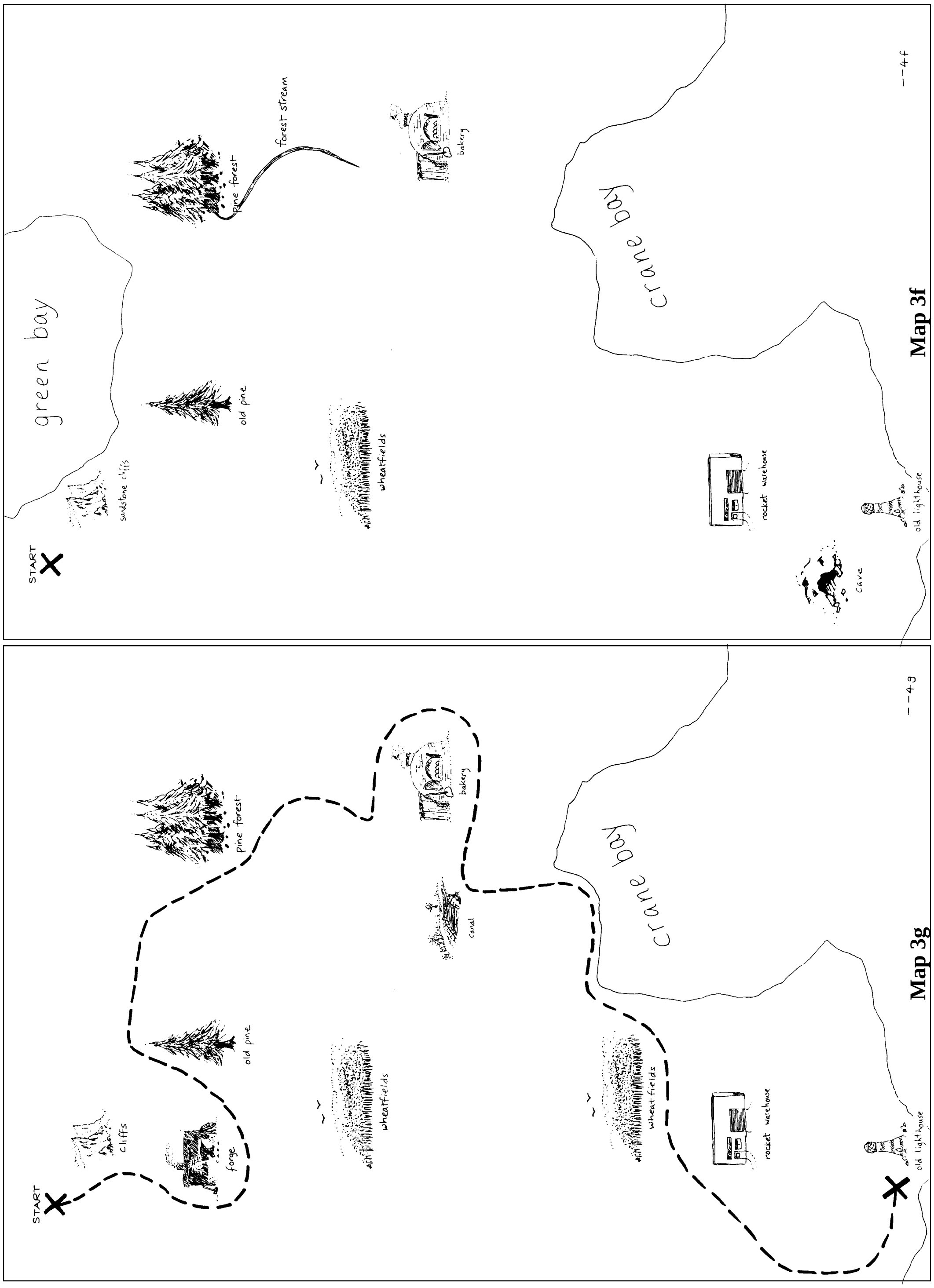

Figure 4.1: Sideways View of Instruction Giver's map (with dotted path) and Instruction Follower's map, from Map Task Corpus, Quad 1 (corresponds to the dialogue in Appendix A).

bay). Both maps have a starting point, but only the instruction giver's map shows the finishing point. The difference between the maps causes various misunderstandings which the participants have to discuss, and this leads to more complex dialogue than a simpler task would.

The maps were designed to encourage certain phonological phenomena. Names of landmarks were designed to realise glottalisation (e.g. in "white mountain"), nasal assimilation (e.g. in "seven beeches"), t-deletion (e.g. "vast meadow"), and d-deletion (e.g. "reclaimed fields").

4.2.4 Subjects

A total of 64 Glasgow University undergraduates (32 female and 32 male) participated in the Corpus. Their ages range from 17 to 30 with a mean of 20. Most subjects have fairly standard Glaswegian or west coast accents.

4.2.5 Design

The Corpus is organised into *quads*, groups of four subjects. Four sets of maps were used by each quad. Each person in the quad played a map game with the other three quad participants.

The Corpus was designed to incorporate a few different conditions related to the ultimate interests of the researchers involved. These are eye gaze, familiarity, and participant role. The Corpus consists of two overall conditions: one which allowed eye contact between participants and one which did not. Within each eye contact condition there are eight quads. Each quad involved two pairs of subjects. The subjects within a pair were friends. Therefore within a quad, any given person was familiar with one other person and not familiar with the other two. This allowed a condition of familiar and unfamiliar pairs when subjects were mixed within the quad. Each person served twice as instruction giver and twice as instruction follower.

A total of 8 conversations are in each quad, 64 conversations in each eye contact condition, and 128 dialogues in the Corpus. Half of the conversations involve familiar pairs and half unfamiliar pairs.

4.2.6 Procedure

The Corpus dialogues were recorded at Glasgow University, by HCRC staff in the Psychology Department. In a small room, the two participants sat at small desks facing each other. On the desks were the two maps. A low barrier separated the participants and their desks. The barrier prevented each participant from seeing the other's map. For the "no eye contact" condition, an additional barrier was set upon the existing one, high enough to prevent eye contact between the two persons.

Each subject wore a small head-mounted microphone through which analog and digital audio (DAT) versions were simultaneously recorded in stereo (one speaker per channel). In addition, two video cameras allowed one-fourth of the Corpus (Quads 3, 4, 7, and 8 in the eye contact condition) to be recorded visually. Videos captured the face of the instruction giver and an angled view of the face of the instruction follower along with his or her upper body and entire map.

4.2.7 Transcription

Each dialogue was meticulously transcribed orthographically to capture the speech verbatim. Transcriptions were marked for false starts, hesitations, repetitions, interruptions, and speech overlap (so that actors might read them as a script). A different transcriber proofread the dialogues, making a second pass through the Corpus. A third round of checking was later performed.

4.2.8 Read Dialogues

One-fourth of the Corpus (Quads 1, 3, 4, and 5 in the "no eye contact" condition) was recorded again, with the original participants reading the carefully prepared orthographic transcripts of their conversations. These dialogues are from the "no eye contact" condition in order to maximise the communicative content carried in the transcript. (A lack of eye contact forces information through the auditory channel. Some information in the eye contact dialogues may be communicated through eye gaze.) The read dialogues involve a reading task and as such are not strictly part of the Map Task Corpus.

The recordings of the read dialogues occurred a few months after the original recordings. Participants did not have the opportunity to rehearse the reading aloud. They were handed transcripts and within a few minutes or less had to re-enact the conversations by reading aloud the dialogue transcripts. All original participants were used except in the case of Quad 4 in which one pair of participants had to be replaced. (Two of the researchers involved in the project took their place.) Other conditions remained the same as in the original recordings, e.g. same maps and same laboratory location[1].

4.2.9 Some Facts about the Corpus

The Corpus consists of 128 dialogues: 8 quads of 8 dialogues each in the two eye contact conditions, making a total of 16 quads. This comprises 20,675 conversational turns, 1,939 word types and 146,855 word tokens. The dialogues total more than 15 hours and average approximately 7 minutes per conversation.

The eye contact condition was found to affect dialogues in a number of ways (Boyle *et al.*, 1994). Subjects with eye contact complete the task more efficiently in terms of information transfer and management of turn taking. Dialogues in the "no eye contact" condition are longer overall, having a greater number of word tokens and turns, though fewer words per turn. The lack of eye contact reduces smooth flow of conversation, as subjects interrupt each other almost twice as often as those with eye contact. (Familiar pairs also tend to interrupt each other more often.) Subjects in the "no eye contact" dialogues overlap their speech more often and produce a greater number of backchannels.

The visual channel may carry some communicative information that is not

[1] Some problems occurred in the recording of some of the read dialogues. The read recordings of Quad 3 mixed both channels, creating a monophonic recording.

present in the auditory channel. For example, eye gaze sometimes appears to provide acknowledgement. Also, instruction followers tend to look up at problem points. Boyle *et al.* note that instruction followers look up significantly more often while discussing features that differ on the maps than while discussing features that do not differ.

Task performance was found (by Boyle *et al.*) not to differ between subjects with eye contact and those without. Performance was measured by comparing the original path to the instruction follower's drawn path, overlaying a centimetre square grid (the scoring system was developed by Anderson, Clark & Mullin, 1991). Scoring counted each grid square between the original path and the follower's path.

4.2.10 Dialogue Excerpt

The following is an excerpt from a Map Task Corpus dialogue. It is the beginning of Quad 1 Conversation 6 in the "no eye contact" condition, between an unfamiliar pair of participants. Speaker A is the instruction giver. (The full dialogue can be found in Appendix A.) This excerpt demonstrates one strategy which subjects may adopt – comparing landmarks on the maps to establish points of common reference immediately before they begin discussing part of the path.

A Okay, the start's at the top left.

B Right, aye, I've got the start marked down.

A You have cliffs there?

B Sandstone cliffs?

A Yeah.

B Mmhmm.

A You don't have a forge, do you?

B No.

A Right, there's a forge about two inches beneath the cliffs. Okay?

B Right, directly down?

A Yeah.

4.3 Conversational Games Analysis

The *Conversational Games* Analysis was developed with reference to task-oriented dialogues from the Map Task Corpus. It was intended to be a system which could represent conversational activity in an independent manner, free from the influence of any potential application. What emerged is a theory of dialogue at two levels.

4.3.1 Functional-Intentional Approach

The main goal in this analysis of dialogue is to represent the function of utterances within a conversation. Function relates to speaker intention in that the speaker intends the utterance to achieve something. Participants achieve goals by eliciting help through verbal means from the other participant. In the case of Map Task dialogues, the goals may be to find out if a particular landmark exists or to get the listener to draw a segment of the path.

The unit of linguistic interaction which accomplishes a goal is the *conversational game*. It is at the level of an exchange (as in Sinclair and Coulthard 1975). Conversational games are composed of *conversational moves*, units at the level of speech acts, which accomplish particular functions. The Games Analysis differs from Sinclair and Coulthard's work in several respects. An important distinction is that Sinclair and Coulthard do not characterise their exchanges in terms of goals and intentions. Also, their analysis does not represent the nesting of interactions – something quite commonly occurring in conversation and particularly in the Map Task Corpus.

4.3.2 Development

Development of the Games Analysis began with an adaptation of Power's (1974, 1979) and Houghton's (1986) work. Power addresses the structure of computer-generated dialogue arising from a task between simulated robots and introduces conversational procedures which handle such interaction. Houghton, building on Power's work, proposes "interaction frames" to represent (and generate) dialogue between two robots which accomplish a simple task involving opening doors and moving through them. His four interaction frames handle attention getting, information giving, information getting, and accomplishing an action. The frames represent the requirements and procedure by which conversational interaction may occur. They define the allowable participants, end goal, effect, preconditions, response, and reply. For instance in the "Making Something Known" frame, the precondition is that the initiator knows that the addressee does not know the information to be shared.

Power (1974) introduces the notion that conversational interaction is a "game". Houghton calls a "game instance" a record of the interaction including the

- type of interaction planned
- identity of the initiator of the interaction
- identity of the addressee
- message which initiated the interaction
- reply received from the addressee
- topic that the interaction involves

The idea of the Conversational Games Analysis is to represent similar sorts of interactions in real human dialogue (not robot dialogue).

Conversational goals are accomplished in games, the unit of interaction which includes the conversational turns necessary to complete a game once it has begun. Moves are the contributions by individual participants in the conversation, identified by their intended function.

4.3.3 Units of Discourse Structure

A game is a theoretical concept in conversational interaction which accomplishes some underlying goal. A move is a component of a game, in which some function is accomplished as an initiation, response, or feedback move. A game minimally consists of an initiation move and a response move. Participants understand implicitly the rules of a game, for instance, that an initiation move expects a response move, and that the responder should contribute appropriately to the game. In this respect, conversational games are similar to other games in which behaviour follows accepted norms.

Games are expected to be well-formed and almost always contain a response move. Otherwise a game is considered to be abandoned. Some games may nest within other games, e.g. as participants realise that more information is needed to accomplish the first game.

Unlike Houghton's analysis in which a generated sentence forms a move, a move in spoken dialogue does not always map to a sentence, or an utterance. Since moves are defined functionally, they encompass portions of speech which accomplish one function.

Conversational games and moves may be characterised as follows:

move An utterance, part of an utterance, or several utterances, which communicate one idea or intent, and serve a particular communicative function. It is uttered by only one person and often ends with a pause. It may continue over two or more conversational turns.

game A series of moves, usually two or more turns, which are necessary to accomplish a conversational goal. One participant initiates a request or exchange of information, for instance, and appropriate responses follow until the interaction is completed. Each participant understands the implicit structure and rules for each game.

A game ends when both participants agree it has ended. As an example, a game might involve one person giving an instruction and end when the listener has understood and has agreed or refused to carry out the instruction.

The basic structure of a game usually consists of two or more moves – an initiating move and one or more response and feedback moves, which may in turn initiate nested games. Nesting often occurs when one of the participants decides that more information needs to be exchanged before the original game can continue.

Game structure follows the natural flow of dialogue. Where one game ends, another begins. Since the Map Task generally involves giving a series of instructions, the dialogues often consist of a series of *Instructing* games, with occasional other games in between.

The structure of a game is not fixed. It depends upon the dialogue context. Some games are short. Others are long, regardless of nesting. Sometimes response and feedback exchanges loop until both participants agree that the initiating move has been satisfied and the game is thus completed.

In developing this system of discourse analysis, many problems presented themselves. Firstly, utterances sometimes appear to accomplish two functions at the same time. (Stubbs, 1983, also finds this to be the case.) For example, an acknowledgement to a move may also serve to hold the turn for the speaker who is then going to give an instruction or ask a question: "*Right* I'd like to you to go down two inches." The Games Analysis does not handle duality of function. It forces a choice in classifying an utterance as one move or another. So the analyst must make a best guess as to the main function of the utterance and make a note that perhaps it accomplishes an additional function. Moves are coded according to the speaker's intention.

The second problem is a more general one, that of deducing the speaker's intention. The function of an utterance is necessarily linked to its intended interpretation. Hence, it is necessary to determine the speaker's intent. Usually this is not a problem. Most people understand each other in conversation. This is why communication flows effectively. However, we do sometimes misunderstand one another or misinterpret particular utterances. Since people cannot read minds, we do not always know what the other person intends with what they say (and it is not always clear that the speaker even knows). Although both speaker and analyst sometimes encounter problems in assessing the intention of utterances, one's best guess is almost always a good guess.

4.3.4 Moves and Games in the Map Task Corpus

Real human dialogue such as that in the Map Task Corpus (Anderson *et al.* 1991) is different and more complex than the one Houghton's robots accomplish, so Houghton's system had to be expanded, although the games could be represented more simply. The repertoire of conversational games was tailored to the dialogues from the Map Task Corpus while keeping in mind application to other dialogues (e.g. Maze Task dialogues introduced below in Section 4.4). The aim was to produce a system of analysis at the level of interaction frames and their components and to identify a set of games and moves which adequately distinguish discourse functions in the Map Task Corpus while not being too specific.

The repertoire of games and their components, moves, increased from an initial four games and eight moves (initiation and response, or opening and closing, from each interaction frame). For example, Houghton's "Getting Information" frame could be separated into at least two games which appear distinct within the map task, for example a *Querying* game and a *Checking* game. The former game occurs when the initiator seeks unknown information, while the latter occurs when the initiator believes the answer is known and wishes confirmation.

Because the Games Analysis was developed from dialogues in the Map Task Corpus, the current repertoire of games and moves reflects the nature of the Map Task. Twelve conversational moves appear in a total of six games. The Map Task generally involves one speaker instructing another, and consequently many of the games in the task are *Instructing* games. The repertoire of games is not intended to be restricted. The working repertoire matches the type of dialogue used for analysis. For Map Task dialogues and other types of task oriented dialogue (such as that elicited by the Maze Task described in Section 4.4) the repertoire introduced below is appropriate, but for other types of dialogue, it could change. For instance, in a telephone dialogue a Greeting game may be needed. One might also decide that a particular dialogue requires plan-related moves such as Power's (1974) SUGGESTPLAN or AGREEPLAN.

Moves which initiate games

The Games Analysis based upon Map Task dialogues has generated six games and twelve moves. Six moves initiate games and six serve as response, feedback, or cue. The six games are *Instructing, Checking, Querying-YN, Querying-W, Explaining*, and *Aligning* games. Their initiating moves are defined as follows (expanded from Kowtko *et al.* 1992):

INSTRUCT This move communicates a direct or indirect request or instruction to be carried out. It contains sufficient detail and clarity for the listener to then act upon the information and do as instructed. The surface form can vary, but the move must serve to instruct the listener. Examples follow (italics are added for purposes of syntax comparison):

"*Go* round, ehm horizontally underneath diamond mine [...]."

"Ehm, *you go* forward from there and you branch off"

"And then when it comes to crane bay *you're keeping* quite close to the coast."

"*You want to go* straight down."

"*You've got to go* up, ehm, to your left again [...]."

"Ehm, could we, ehm, up *we want to go* north through the graveyard and above the carved stones."

"*And I want you to go* towards the left-hand side of the page."

"Ehm, *if you go* under cattle stockade and then loop up slightly [...]."

"*So you're going* down and then along to the trout farm. [...]"

"*Then branch off* above the attractive cliffs [...]."

"Ehm, *round* above horizontally over above the gold mine."

"*And straight down* til about two inches from the bottom of the page."

EXPLAIN This move provides information which the game initiator believes is not yet known by the other person (i.e. new information). The information given is not elicited (that would be a response move). It can relate to anything concerning the task, e.g. landmark existence or action. It may describe the status quo or the position in the task with respect to the goal. Examples follow:

"I don't have a ravine."

"Well I've got a gold mine as well you see."

"I've also got a ravine, below the carved stones."

"I've got that marked as well."

"No green bay."

"So I'll write footbridge."

"I'll just go down here then."

"I'm underneath it now."

ALIGN This move checks that the listener's understanding (which could be in terms of e.g. plans, location in the task, or goal accomplishment) aligns with that of the speaker. It ensures attention, agreement, or readiness. The initiator in this game makes sure that the level of understanding between participants is aligned. This is often realised as a check that both are at the same point in the task or that the responder is ready to go to the next game. The speaker expects the hearer to utter a positive response when the hearer is "aligned" with the speaker. The speaker usually only uses this move when expecting that the hearer will immediately respond positively. A negative reponse indicates a difference of understanding. Examples follow:

"Okay?"

"Do you see what I mean?"

"Shall we begin then?"

“[...] see where it’s written white mountain?”

CHECK Asks a question, the answer to which the initiator believes he or she already knows. (If the initiator lacks information, that would be a query.) The initiator checks self-understanding, usually to see if an instruction was heard or understood correctly by requesting confirmation. Examples follow:

“So you’ve got a diamond mine and a gold mine?”

“So going down to Indian country?”

“And you’ve got the old temple marked?”

“You’re not anywhere near that?”

“You do have a chapel?”

“A straight line between them?”

“Above Indian country?”

“Green bay?” (repeats part of other speaker’s utterance)

“Turning left?” (repeats part of other speaker’s utterance)

“Upwards?”

QUERY-YN This move is a Yes-No question which asks for information previously unknown to the initiator (new information). Examples follow:

“Do you have a trout farm?”

“Have you got the graveyard written down?”

“Have you got poisoned stream marked by the footbridge?”

“Have you circled them? [...]”

“From the abandoned truck?”

“Underneath?”

QUERY-W This move is an open content (e.g. Wh-) or limited choice question which asks for new or unknown information. Examples follow:

“If I ... When I’m by crane bay which direction have I been coming from?”

“Towards what?”

“In where?”

"Above it or below it?"

"What finish?"

Other moves

The following six moves serve as response, feedback, or cue moves within a game:

ACKNOWLEDGE This move indicates vocal acknowledgement of having heard and understood. It is not specifically elicited but often expected before the other speaker will continue, in essence a request to 'please continue'. It announces readiness to hear the next move. It may close a game. Examples follow:

"Okay."

"Ah."

"Oh right."

"Cavalry" (repeats part of other speaker's utterance)

"No?" (repeats other speaker's utterance)

"Above the carved stones, okay." (repeats part of other speaker's utterance)

"I see. Mm that's interesting."

CLARIFY This move clarifies or rephrases given, known, or otherwise old information. Examples follow (same speaker's previous utterance is in angled brackets):

⟨so you want to go [...] actually diagonally so you're underneath the great rock.⟩ "diagonally down to un uh horizontally underneath the great rock."

⟨[...] go to your left of bandit territory and just above it put a cross for finish.⟩ "Ehm a bit well I'd say about eh an inch and a half left from bandit territory just above it."

REPLY-Y This reply move has an affirmative surface form and usually indicates agreement. It is an elicited response (to QUERY-YN, CHECK, or ALIGN). Examples follow (other speaker's previous utterance is in angled brackets):

"Okay."

"Uh-huh"

"I do."

"Oh right. Aye."

"Right okay that's fine."

"Yes got Indian country."

⟨So beneath the great rock?⟩ "Right, beneath the great rock."

⟨Above Indian country?⟩ "Above Indian country" (repeats part of other speaker's utterance)

REPLY-N This reply move has a negative surface form and usually indicates disagreement or denial. It is an elicited response (to QUERY-YN, CHECK, or ALIGN). Examples follow (other speaker's previous utterance is in angled brackets):

"No."

"No, no graveyard."

"No I don't."

"Nowhere near the coast."

⟨Do you have the cavalry?⟩ "No there's no cavalry on this map."

REPLY-W This move is an elicited reply that is not a CLARIFY, REPLY-Y, or REPLY-N. It provides new information. It can be a response to a query that is not easily categorizeable as positive or negative, e.g. "Down." More examples follow (other speaker's previous utterance is in angled brackets):

⟨And across to?⟩ "The pyramid."

⟨Towards where?⟩ "Green bay, at the top."

"At the chestnut tree."

⟨Which direction have I been coming from?⟩ "[...] you've ... you come from vast meadow."

READY This move cues the speaker's intention to begin a new game and focuses attention on the speaker who holds the turn in preparation for the new move. It indicates that the previous game has just been completed, or that the speaker is leaving the previous level or game. Examples follow:

"Okay"

"Right," so we're down past the diamond mine? [...]

The READY move does not achieve the same status of the other moves listed above because its primary function is as a transitional cue, not response or feedback. It almost always precedes an initiating move of a game, and may be placed either at the beginning of a new game (without disqualifying the initial game status of the subsequent move) or between games.

Features

In conjunction with move classes, the Games Analysis includes a set of features which may append to the labels for each conversational move.[2] In the Map Task Corpus, move labels are assigned to each move within a conversational turn. That is, if a move continues over several turns, it has separate labels for each segment, one per turn. Hence, there is a need to mark the move label as being continued (one of the features below).

–aban Abandoned utterance; could be a fragmented sentence. (Abandoned games are marked as such.) E.g.

"Have you got a ..." QUERY-YN–aban

–cont Continuation of previous move, usually after a break or interruption; connects two utterances that should logically be one. E.g.

"And along underneath the diamond mine?" ALIGN–cont

"along to the trout farm. Underneath the trout farm." INSTRUCT–cont

[2] These features are not mentioned in the Kowtko *et al.* (1992) paper but were an original part of the system and have been used in coding the Map Task Corpus.

–interj Interjection; exclamatory phrase which need not connect directly with the topic of dialogue but must not be as remote as (–meta); usually tags onto an ACKNOWLEDGE, EXPLAIN, or INSTRUCT; gives the speaker time to think. E.g.

"Wait a minute." INSTRUCT–interj

"Oh oh!" ACKNOWLEDGE–interj

"What? Okay." ACKNOWLEDGE–interj

–meta Meta-level move, meta-task or meta-linguistic; may involve talking about the time of day, experiment design, room, etc. E.g.

"I've mucked this up completely have I?" CHECK–meta

"I'll start in the right place this time. ..." EXPLAIN–meta

"I sound as if I'm making an awful lot of mistakes here." EXPLAIN–meta

–mumbl Mumbled utterance, whispered; not necessarily intended for other person to understand. E.g.

"I'll need to avoid that" EXPLAIN–mumbl

–repo Move which repeats part or all of other's previous utterance; when on a CHECK or ACKNOWLEDGE it functions to check accurate transmission. E.g. (other speaker's previous utterance is in angled brackets)

⟨Ehm, right and you're turning left up there.⟩ "Turning left?" CHECK–repo

–reps Move which repeats part or all of same speaker's previous utterance usually from another conversational turn. E.g. (same speaker's earlier utterance is in angled brackets)

⟨Do you have the diamond mine?⟩ ...2nd turn later... "Do you d You don't have diamond mine though?" QUERY-YN–reps

⟨Have you got a ...⟩ ...next turn... "Have you got a parched river bed?" QUERY-YN–reps

–fill Move which attempts to complete the other speaker's utterance. E.g. (other speaker's previous utterance in angled brackets)

⟨So you're going down and then⟩ interrupts with "Along to the trout farm." ACKNOWLEDGE–fill

Features serve not as an integral part of the analysis, but to provide additional information to aid different types of analyses. They specify whether a move in a conversational turn is a continuation or is abandoned, mumbled, repeated, interjected, occurs at a meta-conversational level, or serves as a filler. The features which are of particular interest with respect to game structure, are *cont* and *aban*. A labelled move which continues, e.g. INSTRUCT–cont, will never cause a new game to initiate. Likewise, one with an abandon label will end a game prematurely.

4.3.5 Basic Game Structures which Appear in the Map Task Corpus

Although no particular game structure is assumed by the theory of Conversational Games, some basic structures appear in the Map Task Corpus. The structures described below are taken from the analysis of eight Map Task dialogues. They comprise a minimal skeleton structure, the shortest game structure that actually appears.

The sequence of moves in these structures involves alternating speakers, e.g. if the instruction giver initiates the game, the instruction follower responds and the instruction giver may follow up. The equals sign (=) indicates possible choices for a move. Parentheses indicate optional moves. Comments appear in square brackets ([]).

The structures implicitly allow for certain variations. Any move may be interrupted by a –fill attempt, as these types of move occur when one person is attempting to complete the other's phrase. The following example occurs when one speaker interrupts and tries to finish the other's explanation, effectively acknowledging the explanation in the process.

EXPLAIN
ACKNOWLEDGE–fill
EXPLAIN–cont

In this case, the person giving the explanation resumes after the –fill move.

Aligning game

Often, this game is found embedded at the end of an *Instructing* game, when the instruction giver thinks the follower has completed the instruction or is at a point of understanding and just wants to check. Positive response is necessary for the game to work, indicating that communication is successful. Basic structure is

ALIGN

or

ALIGN
REPLY-Y
(ACKNOWLEDGE)

The *Aligning* game which lacks verbal response is one in which the initiator assumes that the lack of a response indicates a positive response. The basic structure occurs in 75% (40 of 53) of *Aligning* games in the sample eight dialogues.

Checking game

This game is usually nested within an *Instructing* game, after the INSTRUCT or subsequent ACKNOWLEDGE move. *Checking* is rarely done at the top-level, but when a game does occur at the top level, it is initiated by the instruction giver, and it has a simple structure of two or three moves. These structures occur in 74% (57 of 77) of *Checking* games.

CHECK
REPLY-Y = REPLY-N [agrees]
(ACKNOWLEDGE)
(ACKNOWLEDGE)

or

CHECK

REPLY-N = CLARIFY

ACKNOWLEDGE

In the latter construction, the qualified response is followed by an ACKNOWLEDGE move which signals that the change in information has been understood.

Explaining game

This game occurs most often as a nested sub-game within an *Instructing* game, but also occurs at the top-level by both participants.

EXPLAIN

or

EXPLAIN

ACKNOWLEDGE

These structures occur in 79% (45 of 57) of games.

Instructing game

Because the instructions describing the route are often complicated, the structure of *Instructing* games can be lengthy. Other games may embed inside the *Instructing* game. *Instructing* rarely occurs as a nested game. The following four structures are those which appear most often. They account for 50% (70 of 141) of games.

INSTRUCT

ACKNOWLEDGE

(ACKNOWLEDGE)

or

INSTRUCT

ACKNOWLEDGE

INSTRUCT cont

ACKNOWLEDGE

(*the last two moves repeated up to 3 additional times*)

or

> INSTRUCT
>
> *embedded game*

or

> INSTRUCT
>
> ACKNOWLEDGE
>
> *embedded game*

Querying-YN and Querying-W games

These games occur at the top-level and in nested position. They may have a *Checking* or, less commonly an *Instructing* or *Explaining* game nested after the initial move or second move. Most games are short.

> QUERY-YN
>
> REPLY-Y = REPLY-N = REPLY-W
>
> (ACKNOWLEDGE)

and

> QUERY-W
>
> REPLY-W
>
> (ACKNOWLEDGE)

The structures occur in 60% (57 of 95) of *Querying-YN* games and 44% (12 of 27) of *Querying-W* games.

4.3.6 Examples of Games in Dialogue

Map Task dialogues typically start with a series of *Instructing* games, although a small number begin with *Explaining* games in which each participant compares landmarks on the maps. The following excerpt from dialogue NAQ3C8 shows one large *Instructing* game with three nested games: *Explaining*, *Querying-YN*, and *Checking*. The instruction follower completes the instruction by the end of the embedded *Checking* game. Angled brackets indicate overlapped speech, and vertical lines indicate the boundary of a move:

A Right,| em, go to your right towards the carpenter's house.

READY | INSTRUCT

B All right

ACKNOWLEDGE

well I'll need to go below. I've got a blacksmith marked.

EXPLAIN

A Right, well you do that.

ACKNOWLEDGE

B Do you want it to go below the carpenter?

QUERY-YN

A ⟨ No, | I want you to go up the left hand side of it towards /

REPLY-N | INSTRUCT–cont

B Okay.

ACKNOWLEDGE

A Green Bay and make it a slightly diagonal line, towards, em sloping to the right. ⟩

INSTRUCT–cont

B So you want me to go above the carpenter?

CHECK

A Uh-huh.

REPLY-Y

B Right.

ACKNOWLEDGE

In the above excerpt, it is not clear whether the final ACKNOWLEDGE ends only the larger *Instructing* game. Therefore, it is coded as ending both the nested *Checking* game and the larger *Instructing* game simultaneously.

The next example shows that the *Aligning* game occurs at the point where the instruction giver thinks the requested action has been completed and it is safe to ask if the hearer's path and location in the dialogue are up to date. The initiator expects an affirmative response.

B From the top of the white mountain?

CHECK

A From the top of the white mountain. Right up.

REPLY-Y

B How f-... How far up?

QUERY-W

A ⟨Ehm, about six centimetres till you're about, ah, till you're about ehm, about the same distance away from... you're about five centimetres... six centimetres from the top of the page /

REPLY-W

B Right. Okay.

ACKNOWLEDGE

A now. ⟩

REPLY-W–cont

B Okay.

ACKNOWLEDGE–cont

A Right,| you're to the left hand side of the page.

READY | ALIGN

B Yes.

REPLY-Y

4.3.7 Statistics from the Map Task Corpus

A total of 8,899 games appear in the 128 Map Task dialogues, making an average of 70 games per dialogue. Table 4.1 shows the percentage of each type of game and the percentage of those games which are (1) embedded and (2) spoken by the instruction giver. The most common game (by a small margin) is the *Instructing* game. It generally occurs at the top level and is initiated by the Instruction Giver. In contrast, the *Checking* game rarely appears at top level. It is almost always embedded inside another game and is usually initiated by the Instruction Follower.

Table 4.1: Proportion of the 8,899 Games Appearing in the Map Task Corpus (MTC), Proportion which are Embedded, and Proportion Initiated by Instruction Giver

Game	*MTC %*	*% Embedded*	*% by Giver*
INSTRUCT	22	8	99
ALIGN	17	83	93
QUERY-YN	17	56	67
QUERY-W	8	81	29
CHECK	20	95	15
EXPLAIN	16	73	41

Embedding occurs up to 4 levels deep (plus the top level) in the Corpus, but this depth does not occur often as it makes conversational progression very difficult. Most embedding occurs up to one or two levels deep.

A total of 25,945 moves appear in the Corpus, making an average of 203 moves per dialogue. Here, moves are counted per conversational turn. That is, a move which continues over two or more conversational turns is counted separately in the different turns. More than one type of move may occur in any given conversational turn (as one can see by comparing 25,945 moves with 20,675 turns in the Corpus). Table 4.2 shows the percentage of moves in the corpus and what percentage of each move has a feature. The "cont" feature indicates the moves which are continuations across conversational turns. Table 4.3 shows the statistics for features in the corpus. Some moves may have more than one feature attached. Note that coding of moves in the Map Task Corpus was very strict with regard to continued (–cont) and abandoned (–aban) moves, and less strict with regard to the other six move features.

The most common move is ACKNOWLEDGE. Second is INSTRUCT. These frequencies link to the nature of the Map Task in which instructions are given by one participant as the other listens, providing feedback and acknowledgement as necesary. Approximately one-fourth of the moves consist of some type of question (ALIGN, QUERY-YN, QUERY-W, and CHECK). Almost one-fifth of the moves are REPLY moves. The frequency of these moves indicates that the dialogues involve a significant amount of questioning as well as instructing. The obstacles that subjects encounter in the Map Task cause them overall to spend much of their time enquiring about particular landmarks and detailed directions of the route.

Approximately one-half of the INSTRUCT moves continue across conversational turns. One-third of CLARIFY moves, and one-fourth of EXPLAIN and REPLY-W moves also continue across turns. Thus, if one looks at moves including their continuations, the percentages change slightly. The number of whole moves which occur in the Corpus is actually 20,772. The percentages of whole moves within the Corpus for the respective moves INSTRUCT through CLARIFY, as in Table 4.2, would then be 10%, 7%, 7%, 3%, 8%, 8%, 14%, 4%, 3%, 24%, 9%, and 4%. The main difference in counting moves in this manner is that ACKNOWLEDGE moves now far outnumber any of the other move categories, providing almost one-fourth of the inventory.

Table 4.2: Proportion of the 25,945 Moves (as counted per turn) in the Map Task Corpus (MTC) and Proportion which Have Features: Continued Move, Abandoned, Filler, Interjection, Meta-Level, Mumbled, Repeating Other Speaker, and Repeating Self ("-" indicates negligible)

Move	*MTC %*	*% C*	*% A*	*% F*	*% I*	*% Me*	*% Mu*	*% RO*	*% RS*
INSTRUCT	16	53	3		1	-		-	-
ALIGN	7	16	-		-				-
QUERY-YN	7	13	2	-	-	1		-	-
QUERY-W	3	13	3		4	1		1	1
CHECK	8	20	2	1	1	-		8	1
EXPLAIN	8	24	3	-	3	6	-	-	-
REPLY-Y	12	9	-	-	1	-		9	-
REPLY-N	3	11	-	-	1	-		3	1
REPLY-W	4	25	3		2	1		3	3
ACKNOWLEDGE	20	4	-	-	2		-	8	-
READY	7	2			2				-
CLARIFY	5	31	1	-	1			3	3

4.4 Coding Agreement Experiment: Naïve Subjects

4.4.1 Purpose

Coding is of little use for empirical work unless analysts can replicate it. In order to test whether naïve coders could learn the system quickly, an experiment was conducted (as detailed in Kowtko, Isard and Doherty-Sneddon, 1992).[3] The

[3] As the first author, I was the primary developer of the Conversational Games Analysis and wrote instructions for the coding agreement experiment described here, while the third

Table 4.3: Proportion of the 25,945 Moves (as counted per turn) in the Map Task Corpus (MTC) which Have the Following Features: Continued Move, Abandoned, Filler, Interjection, Meta-Level, Mumbled, Repeating Other Speaker, and Repeating Self ("-" indicates negligible)

Feature	*% of Moves in MTC*
–cont	19
–aban	1
–fill	-
–interj	2
–meta	1
–mumbl	-
–repo	4
–reps	1

experiment compared the coding of naïve subjects to that of experts in their labelling of conversational moves in two different types of task-oriented dialogue. It also tested the ability of naïve coders to determine move boundaries, by assessing agreement scores between the two groups.

Although the coding system was developed using Map Task dialogues, it seemed an appropriate opportunity to test the system's robustness by employing dialogues arising from another task, the Maze Task (described below).

4.4.2 Subjects

Four third-year undergraduates in the Psychology course at Glasgow University agreed to serve as subjects, as part of their course. The group consisted of two female and two male students. They had never before performed an analysis of discourse and were not familiar with the relevant techniques or literature.

author actually conducted the experiment and personally instructed the subjects.

4.4.3 Materials

Two types of task-oriented dialogue formed the basic experimental material: two Map Task dialogues (from the Corpus; See Section 4.2) and three Maze Task dialogues (described below). Both transcripts and audio tapes of the conversations were provided. Expert coders (Doherty-Sneddon for the maze dialogues and myself for the map dialogues) had marked conversational move boundaries on each of the five transcripts presented to the subjects. The subject, or naïve coder, then had only to assign move labels to each portion of transcript, not determine move boundaries. The tasks for the subjects were deemed appropriate considering that they participated as part of their third year projects, lasting one academic term. It seemed more important to test the subjects on move distinctions but not require them explicitly to learn how to determine move boundaries.

Nevertheless, the ability of subjects to find move boundaries was considered potentially interesting. An additional (third) unmarked Map Task dialogue (also in transcript and audio version) was provided so that subjects could specify move boundaries.

Other materials consisted of a detailed set of instructions for the subjects. These instructions described the Conversational Games Analysis and the general goal of the experiment. They explained the concept of a game and its conversational moves and provided example game structures based upon three Map Task dialogues (an earlier version of the information in Sections 4.3.4 and 4.3.6).

Maze Task

Maze Task dialogues were taken from a corpus of dialogues between children (of late primary school age). Garrod and Anderson (1987: 185-187) describe the Maze Task as follows:

> The maze game was designed to elicit natural dialogues containing spontaneously generated descriptions of locations within a prede-

fined spatial network, where the exact positions described could be independently verified by the experimenter.

The essence of the game was as follows. Each player was seated in a different room confronted with a VDU on which a maze was displayed. The mazes consisted of small box-like structures connected by paths along which the players could move position markers [...].

The purpose of the game was for the players to move the position markers through the maze (one path link at a time) until they had both reached their respective goal positions. Furthermore, each player could only see his own start position, goal and current position marker.

The co-operative nature of the game arises from two additional features of the mazes. First each maze contained obstacles in the form of gates which blocked movement along the paths where they were positioned [...]. Secondly, there were certain nodes which were marked as switch positions and, like the gates, these were distributed differently for each player. It was in overcoming the obstacles that verbal co-operation was required, since the fundamental principle of the game was as follows. If a given player (say A) moved into one of the switches marked on the other's (B's) screen, then the entire configuration of B's gates would change. All paths that were previously gated would be opened and all those previously open would be gated. Therefore when a player required the gates to be changed, they would have to enlist the co-operation of the other player, find out where he was located and then guide him into a switch node only visible on their own screen.

[...]

Typically, a game would therefore consist in players attempting to move towards their respective goals with dialogue intervening between moves. The dialogue would contain descriptions of the play-

> ers' current positions in the maze, switch node locations and goal positions with each speaker's contribution to the dialogue recorded on a separated channel [...]

The Maze Task, like the Map Task, was designed to elicit conversation and co-operation between the participants, as success in the game requires co-operation. To win the maze game, both players must reach their finish points within 10 minutes. No role structure is imposed upon the participants, such as that of instruction giver and follower in the Map Task. Because of the co-operative nature of the task, players negotiate their own roles.

4.4.4 Procedure

Subjects were first given tutorials on the topic of conversation analysis by the experimenter (Doherty-Sneddon). The purpose was to provide a basic understanding of how to analyse discourse since the subjects had never done this before. Background reading consisting of several articles was distributed and discussed, to familiarise them with general theory and methods of discourse analysis.

While learning about discourse analysis, the subjects were taught some of the basics of the Games Analysis and shown a detailed example of a coded Maze Task dialogue.

After the introduction to discourse analysis, subjects were handed the materials necessary for the experiment: transcripts and audio tapes of the six conversations, and the set of written instructions describing the move categories. They were asked to read through the instructions and assign a single move label to each move as marked in five of the transcripts (of two Map and three Maze Task dialogues), writing the label name on the copy of the transcript. Although some conversational moves seemed to be ambiguous in function, subjects were asked to assign only one label per move, their best choice. After beginning work on the move labelling task, they were asked to assign move boundaries to the sixth dialogue.

Coding the transcripts took several weeks. During that time, the subjects occasionally consulted the experimenter who then clarified instructions, explained particular concepts in detail, or helped with other problems. Each subject coded two Map Task and three Maze Task dialogues and segmented one Map Task dialogue with respect to move boundaries.

4.4.5 Results

In the process of scoring the data, the experimenter discovered that one of the subjects had altered some labels after seeing the expert's coding. His data was therefore not considered in the results.

The Table below (Table 1, in Kowtko *et al.*, 1992: 9) shows the percentage of agreement between each of the three remaining naïve coders and the expert coder in classifying conversational moves. The mean percentage agreement in move classification for the two Map Task dialogues is 82% and the mean for the three Maze Task dialogues is 75%. The mean agreement across the five dialogues (1561 matches, $n = 2018$) is 77%. Agreement scores for both dialogues are significantly different from chance, according to a chi square test ($p < .001, \chi^2 = 239.30, d.f. = 8$). The difference between the scores of the Map Task and Maze Task dialogues is not significant (in a comparison of the scores of one type of dialogue to the mean of the other's scores, $p < .10, \chi^2 = 5.32, d.f. = 2; p < .10, \chi^2 = 8.56, d.f. = 4$; these numbers and statistical test results from Doherty-Sneddon, p.c.).

Results of the assignment of conversational move boundaries are shown in Table 4.5. The mean percentage of move boundary agreement between naïve and expert coder is 88%. Naïve coders both disagreed with placement and added more move boundaries than expert coders had.

Every occurrence of disagreement, or mismatch, between the naïve and expert coders was examined. Although many mismatches are due to the difficulty of guessing the speaker's intention in a dialogue and assigning the best move label to an ambiguous utterance (and thus are a problem inherent in the coding

Table 4.4: Percentage of Agreement between Expert and Novice (Subject) in Classifying Moves from Five Dialogues

Dialogue:	Map 1	Map 2	Maze 1	Maze 2	Maze 3
Subject					
1	84	86	77	70	75
2	86	81	72	73	85
3	81	77	72	69	77

Table 4.5: Percentage of Move Boundary Agreement between Expert and Novice (Subject) from One Dialogue

Subject	Map 3
1	88
2	89
3	86

system), certain types of move mismatch show regular, predictable patterns, revealing misunderstanding of the coding instructions by the subjects. These types of mismatch were named *retrainable* because they reflect a misunderstanding which should disappear when subjects are given more instruction and more time to code. They account for 44% of the move disagreement.

An analysis of the retrainable mismatches occurring between subjects and experts is revealed in Table 4.6 (from Kowtko *et al.*, 1992, p.10). The cause of each mismatch instance in these examples is clear from examining the actual text. A brief description of the cause appears in the last column. The pairs are listed from highest frequency mismatch at the top to the least at the bottom.

The most common retrainable mismatch occurs when a subject codes a REPLY-Y move as an ACKNOWLEDGE move, losing the distinction between an elicited and unelicited utterance. This is a problem of ignoring utterance context. A less common mismatch, but one in which the cause is clearly seen, is

Table 4.6: Plausible Cause for Each Retrainable Mismatch Type

Mismatch Type		Plausible Cause
Expert	*Novice*	
REPLY-Y	ACKNOWL'	Novice misinterpreted an elicited response as an unelicited response.
REPLY-W	EXPLAIN	Novice misinterpreted an elicited response as an unelicited information-giving move.
CHECK	QUERY-YN	Novice has not realised that an interrogative was about known information or does not realise that this is the distinction between QUERY-YN and CHECK.
ACKNOWL'	CLARIFY	Novice has misunderstood the definition of CLARIFY. The function of clarifying or rephrasing has been confused with an acknowledgement of information in which the phrase is repeated.
CHECK	CLARIFY	Novice has misinterpreted a question about information from another speaker as a clarification of that information. A speaker only clarifies information they have previously given.
QUERY-YN	QUERY-W	Novice has not realised that a question requires only affirmation or negation, or does not appreciate the distinction between this and a more open content ('wh') question. Often the syntactic form of a question can cause this confusion.
INSTRUCT	EXPLAIN	Novice has not realised that the function of a move was to instruct a co-participant in an action. Again, the form of an utterance is confused with its function; it looks like an offering of information rather than an instruction.
REPLY-Y	CLARIFY	Novice misinterpreted an affirmative response to a question as a clarification of previous information. It would only be a clarification if the affirmative response followed a question about a message given by the clarifier.
ALIGN	QUERY-YN	Novice was again misled by form. ALIGN describes moves which have the function of obtaining feedback about a listener's knowledge or accomplishment of some task. QUERY-YN has the function of eliciting new information. The novice has not appreciated this distinction.
INSTRUCT	QUERY-YN	Novice has overlooked the INSTRUCT function of a move because its form was interrogative.

the coding of an INSTRUCT move as an EXPLAIN move. This mismatch happens when the coder pays more attention to form than function. Some moves often associate with a certain grammatical structure or type. For example, EXPLAIN moves are often realized by declarative sentences. With more practice, the coder should learn to distinguish function from form more easily. Here is an example of an INSTRUCT move being coded as an EXPLAIN:

G Right. | well, move up and round and above them.
READY | INSTRUCT

F Ok.
ACKNOWLEDGE

G Right. | Across, not as far as a wood. | Do you have a wood?
READY | INSTRUCT (*EXPLAIN*) | QUERY-YN

There will always be some ambiguity as to the intent of certain moves because we cannot read the speaker's mind. The following excerpt illustrates a difficult choice of move class. It is not clear whether the instruction giver believes that the diamond and gold mines are the same or whether the follower has both mines, and the move remains ambiguous between QUERY-YN and CHECK:

G Do you have the diamond mine?
QUERY-YN

F Yes I've got a gold mine.
REPLY-Y

G Ah.
ACKNOWLEDGE

F Hm.
ACKNOWLEDGE

G Do you d- You don't have diamond mine though?
QUERY-YN (*CHECK*)

Some mismatches were clearly errors on the part of the novice coders. Here is an example in which the coder assigned a function EXPLAIN which the move could not reasonably be taken to perform:

G Go diagonally down to un- uh- horizontally underneath the great rock.
INSTRUCT

F Okay.
ACKNOWLEDGE

G Okay. | I-. . .
READY

F So beneath the great rock?
CHECK (*EXPLAIN*)

4.4.6 Discussion and Conclusion

The naïve coders achieve respectable levels of coding agreement with the expert coders. We estimate that with more time for instruction and practice, the formerly novice coders would improve, and the overall agreement in move classification would rise from 77% to 87%.

In addition to being able to label conversational moves, results of the boundary placement experiment (see Table 4.5) show that naïve users of the system are able to segment moves 88% of the time. Although subjects were not specifically taught the criteria involved in determining move boundaries, they were successful in completing this task.

The coding experiment shows that the conversational moves analysis is learnable. The repertoire of moves and the concept of a move is something that naïve users of the system can easily learn with a reasonable degree of agreement with expert users. Also, the system can apply to different types of task-oriented dialogue.

4.5 Coding Agreement Experiment: Expert Subjects

4.5.1 The Experiment

Carletta *et al.* (forthcoming)[4] describes a similar coding agreement experiment between expert coders. In this experiment, four subjects who have extensively used the Games Analysis with the Map Task Corpus have labelled four Corpus dialogue transcripts, given written and audio versions. Each subject marked their own move boundaries, labelled the moves, and marked and labelled game beginnings and endings.

4.5.2 A Summary of Results

Pairwise agreement on the location of move boundaries (where any coder marked a boundary) was 89% ($N = 796$). Agreement on move classification was assessed for moves whose boundaries were agreed. Results are given in terms of the kappa coefficient.[5] Agreement on move labelling was good ($K = .83, N = 563, k = 4$). The most confusion occurred between the following moves: CHECK and QUERY-YN, INSTRUCT and CLARIFY, and ACKNOWLEDGE, READY, and REPLY-Y. When moves are considered in terms of two categories, initiation and response/READY moves, agreement was also good ($K = .89$). Subjects successfully distinguished broad categories of game-initiating move function: commands (INSTRUCT), statements (EXPLAIN), and questions (QUERY-YN, QUERY-W, CHECK, and ALIGN) ($K = .95, N = 243, k = 4$). Agreement was high on classification of the questions ($K = .82, N = 98, k = 4$). Coders agreed well on the five response move categories ($K = .86, N = 236, k = 4$). They exhibited less agreement distinguishing the information-giving moves INSTRUCT,

[4] My role in this research was as primary developer of the Conversational Games Analysis, one of the four subjects in the experiments, and as an author, participant in editing the drafts.

[5] For this application of kappa, Krippendorff (1980) specifies that $K > .8$ indicates a good result and $.67 < K < .8$ indicates a possibly good or borderline result.

EXPLAIN, and CLARIFY ($K = .75, N = 132, k = 4$).

Agreement on move labels was tested with a conversation from a different domain. Two of the expert coders segmented and labelled moves in a conversation between a hi-fi sales assistant and a married couple who wanted to buy an amplifier. Agreement on move segmentation was good ($K = .95, N = 819, k = 2$). Agreement on move classification was also good ($K = .81, N = 80, k = 2$).

Agreement on game labelling (for the Map Task Corpus dialogues) was also checked. Pairwise agreement on where a game begins was 70% ($N = 203$). When subjects agreed where a game begins, they agreed well on the type of game ($K = .86, N = 154, k = 4$). They did not agree very often on whether a game was embedded or not ($K = .46$). There was 65% pairwise agreement on where games ended. Where a game ends partly depends upon whether or not it is embedded, so the confusion on embedding apparently carried through to the location of the end of games.

The most experienced coder coded one dialogue a second time, after a period of two months. Agreement between codings was 90% ($N = 49$) concerning where a game began. When the location of a game beginning agreed, the type of game had good agreement ($K = .88, N = 44$). Whether or not games embedded also showed agreement ($K = .95$). Game endings agreed 89% of the time.

4.5.3 Conclusion

The experiment in Carletta *et al.* shows that coding agreement between expert subjects is very good overall. Although it is impossible to do a direct comparison between the expert–expert agreement and naïve–expert agreement experiment results (because the experts segmented moves themselves), the kappa coefficients indicate that levels of agreement on move labels are high in both cases.

4.6 Using the Conversational Games Analysis

The Conversational Games Analysis makes it possible to describe an utterance in terms of its function at different levels. For example, the simplest level identifies the move. Game context further specifies the function by shaping the utterance's interpretation. The Games Analysis allows an utterance to be referred to as move x within game y. Another level of specification might involve identifying the previous move.

A system of dialogue analysis such as this provides the framework for solving various problems in discourse. In the case of computer dialogue systems, a model of games and moves could represent each user's dialogue state. Such a model could show the user's beliefs in terms of which game and move are current. Because the rules of a conversational game narrow the choice of the next move, each user knows which, of a small selection of moves, to expect to hear and interpret. The analysis also facilitates the study of the function of intonational tune (Chapters 6 and 7), given that the tune reflects an utterance's conversational role.

Chapter 5

Analysis of Intonation

This chapter describes a method of analysing intonation. It complements the previous chapter which introduces the Conversational Games Analysis—an independently designed method of discourse analysis. The two components are equally important in studying the function of intonation in dialogue.

The sections below discuss the choice and implementation of a new system of intonation transcription. The new analysis marks a compromise between the extremes of an acoustic-phonetic analysis and a phonological analysis. It remains relatively close to the phonetic surface form so as not to discard distinctions between contours which may be important. It is simple because the data to which it is applied consists of single words. A simple system is, however, not necessarily easy to implement. Transcription problems are addressed in Section 5.2.2.

5.1 Reasons for Choosing a New Analysis

The studies in Chapters 6 and 7 examine single-word utterances (from the Map Task Corpus explained in Section 4.2) as a means of limiting the complications that longer phrases pose. These brief utterances present their own problems to existing descriptions of intonation, however. Firstly, the utterances contain only one pitch accent, making distinctions such as nucleus location irrelevant. Secondly, the pitch often does not move much, making representation of level tunes an important requirement.

The systems (reviewed in Chapter 3) which appear most able to handle the task of describing the intonation of single-word phrases largely apply to longer utterances in which main prominence is an issue, and distinguishing accents is important. Level tunes in these systems are not well handled.

Consider Pierrehumbert's phonology (1980; also Beckman and Pierrehumbert, 1986), and the ToBI system intonation tier which is based upon it. The rules which link the tonal representation (H and L) with acoustic realisation preclude an unambiguous analysis of low level.[1] The phonology has no means of distinguishing a low fall (L*L-L%) from a low level tune.

The analyses offered by de Pijper (1983) and Taylor (1992, and Taylor and Black, 1994) lack a clear means of representing level tunes on pitch accents. They are more concerned with representing the pitch movement that associates with prominent syllables. Although it is possible to represent a level tune as a connection in Taylor's model, this is not ideal because doing so identifies a pitch accent with a segment (a connection) which usually occurs between pitch accents, and it forces an analysis in which all tunes are represented by numerical parameters, not categories. Categorical distinctions are needed.

The British approach (e.g. Crystal, 1969; O'Connor and Arnold, 1973) offers a single level tone amongst the other possible nuclear tones. The main problem with this approach is that it represents non-level tunes in terms of pitch movement (F0 changes). An analysis using accent target levels is preferred, as it captures the acoustic data better than F0 changes do. (See Section 3.2.2 for details.)

The new system uses target levels to represent the different intonation contours found in single-word utterances in the Map Task Corpus.

[1] Hockey (1991) finds that single-word tunes (with level, rising, *and* falling pitch accents) cannot fit unambiguously into Pierrehumbert's framework. Thus, Hockey rejects the framework and provides a more acoustic description of these tunes.

5.2 Creating a New Analysis

A phonology represents intonation contours at some level of abstraction. The goal for the new analysis was, as a phonology, to represent intonation contours and distinguish them in such a way as to identify the minimal units. Since the utterances in the data are brief, comprised of single words with single pitch accents, common sense suggested that the representation be simple. The brevity of the utterances also suggested that the description might need to remain relatively close to the surface phonetic form (i.e. the acoustic represenation) of the utterances so as not to inadvertently miss important distinctions. The present work uses a simple target representation of pitch movement based upon auditory perception and instrumental analysis.

One can distinguish between two approaches to representing intonation (from Pierrehumbert, 1980: 31):

> One approach attacks the problem by attempting to deduce a system of phonological representation for intonation from observed features of F0 contours. After constructing such a system, the next step is to compare the usage of F0 patterns which are phonologically distinct. The contrasting approach is to begin by identifying intonation patterns which seem to convey the same or different nuances. The second step is to construct a phonology which gives the same underlying representation to contours with the same meaning, and different representations to contours with different meanings.

As Pierrehumbert's phonology does, the present analysis takes the first approach. It addresses the first step in the present chapter and the second step later (Chapters 6 and 7). The Dutch system (developed at IPO) also takes this approach (from 't Hart *et al.*, 1990: 5):

> A language user's intonational competence not only comprises knowledge about melodic form, but also about melodic function. However, the assessment of the formal properties of intonation takes logical

> precedence over the study of its linguistic and expressive use. Eventually we want to come to grips with the communicative value of intonation, but our immediate concern is to develop a descriptive framework for the melodic properties of speech and for the intonational features of language.

The present thesis considers the question of the meaning of intonation categories a separate issue from the construction of a phonology. Intonational meaning is discussed in the studies in Chapters 6 and 7.

5.2.1 Interpretation Issues

Various issues arise in the analysis of brief phrases. Gussenhoven and Rietveld (1991: 425) notice one potential problem with single-accent utterances, suggesting that without a context in which to judge the linguistic interpretation of the accent, the judgement might degenerate into a comparison of "linguistically uninterpreted surface forms". Considering the frequency with which such utterances appear in task-oriented dialogue, the variety of intonation patterns which appear in different discourse contexts (see results in Section 6.4), and the success with which speakers communicate using these single-accent utterances, it seems reasonable to discount such a worry. We can assume that linguistic interpretation is indeed being carried out on single-word utterances extracted from dialogue.

5.2.2 Transcription Issues

In developing an analysis, different issues of implementation arise. These involve the precise nature of the mapping between acoustic representation, e.g. pitch trace, and phonological representation.

There are really two issues at hand. One relates to how we detect prominence. The mapping between prominent syllables and acoustic features is quite difficult to describe in a formal or computational manner (although many have started to work on the problem recently). Certain features relate to prominence, such as high intensity and duration of a syllable, as well as pitch (see discussion in Section 3.1. Of the various approaches to intonation description mentioned in Chapter 3, the Dutch system comes the closest to bridging this gap, as distinctions between types of pitch movement (presumably on pitch accents) are motivated by acoustic data. What makes the problem more difficult is the fact that even experts, i.e. intonation phonologists, have problems identifying prominent syllables. They also sometimes disagree about the location and type of pitch accent in utterances. A great deal of training is necessary to identify accents and boundaries in the acoustic signal. Such training allows phonologists to achieve a more consistent and reliable transcription of intonation.[2] The issue of prominence is reviewed in Chapter 3 and is beyond the scope of this thesis.

Another issue involves more practical matters which relate to using instrumental analysis to corroborate auditory analysis. It includes solving problems which arise from using pitch traces, e.g. the correctness of a trace and how a trace can help (or hinder) the classification of tunes.

This section seeks to identify not so much a set of hard and fast computational rules for mapping F0 points to phonological labels as much as a set of careful guidelines for transcription. Solving the problems in general involves understanding instrumental analysis.

Instrumental analysis reveals a variety of problems in implementing any intonational phonology. This is because a variety of factors influence the resulting F0 values. Phonetic segments can affect computation of the fundamental frequency, introducing errors. A reliable method of analysing the acoustic signal is needed. Perturbation of the F0 trace occurs at different points, such as that

[2] Silverman *et al.* (1992) describe an experiment involving ToBI coding in which the 20 participants have experienced a great deal of training. Pairs of participants agree that a word has accent 83% of the time. When coders agree that an accent is present, they agree 61% of the time on the type or tune of that accent. Participants in another ToBI study (Pitrelli *et al.*, 1994) were also given lengthy instruction, often indirectly, regarding the mapping between levels of F0 and phonological units. Pairwise agreement between 26 transcribers is 80.6% that a pitch accent occurs. When coders agree than an accent is present, they agree 64.1% of the time on the type of accent.

of frication (e.g. on the edge of a vowel and fricative when F0 rises dramatically) and glottalisation. When the voice is has a breathy quality and voiced segments such as vowels are low in volume, (i.e. low voicing probability is computed) pitch periods which lend to the fundamental frequency may be obscured and a pitch trace produces incorrect F0 values. Also, 'paralinguistic' factors such as emotional excitement may cause laughter or other breathing and pitch irregularities which introduces noise into an F0 trace. Problems such as octave errors in the computation of F0 are relatively easy to correct.

It is necessary to set up criteria for distinguishing intonation categories. Ambiguity regarding the presence of pitch movement may suggest one solution when using perception alone and a different solution when consulting the pitch trace. Certain accented syllables with a particular interval of pitch change, say 7 Hertz (Hz), are perceived as containing distinct movement while others with the same interval are heard to maintain their pitch. One can resolve the problem by arguing that perception holds priority over purely instrumental analysis since human speech involves perception, not instrumental computation. (After all, instruments do not detect perceptual categories, people do.)[3] The interpretation of pitch change lending to the presence of movement for data in this thesis depends partly upon assessment of local pitch range and partly upon how well the contour matches other contours in the relevant category.

The concept of pitch range presents its own problems. A speaker's pitch range may change in a relatively short span of time. This makes the task of identifying the relative height of a point in the speaker's pitch range more difficult. For purposes of this study, immediately proximate pitch range is used. The range of the same speaker's nearest utterances are used to assess the local range.

[3] Schubiger (1958) warns the analyst that much of the information available from instrumental analysis is linguistically irrelevant. She suggests discretion and recommends that the analysis be used in corroboration. The present work uses it as a secondary analysis because people communicate effectively without the aid of instruments. Auditory analysis is given primary importance.

5.3 The New Analysis

The present work uses a simple representation of pitch movement based upon auditory perception and instrumental analysis (although transcription may be attempted purely by auditory perception) to describe single-word utterances. It was developed by examining single-word phrases from the HCRC Map Task Corpus (Section 4.2). The Corpus contains Scottish English (the Glasgow accent in particular; see the discussion in Section 3.3). The analyst assigns the pitch accent in each phrase one or more accent targets, high (H) or low (L), associated with a maximum point (e.g. peak) or minimum point (e.g. valley) in fundamental frequency (F0) as detected in the acoustic representation.

5.3.1 Contours

Five combinations of high and low accent targets comprise the set of intonation contours.[4] Each utterance contains one pitch accent and is therefore represented with one of the following contour labels (which can be interpreted in terms of targets or pitch movement):

Label	*Tune*	*Brief description*
H	High level	Level tune high in the speaker's range
L	Low level	Level tune low in the speaker's range
HL	Fall	Higher tone followed by lower tone (simple fall)
LH	Rise	Lower tone followed by higher tone (simple rise)
LHL	Rise-fall	Distinct LH followed by a low tone

The lack of HLH (fall-rise) found in the Corpus accounts for the lack of a symmetrical complement to the complex tune LHL in the table above. Although HLH may exist in a Glaswegian's inventory, it is not included here because the

[4] The analysis began with three target levels, High, Medium, and Low. Medium was found not to be significant and was removed from the final analysis. The instances in which it appeared were reassigned High or Low labels except in the case of the Glaswegian rise, LH(M), in which it is deleted. This is discussed below.

data do not justify its inclusion. The data also do not support the inclusion of more complex tunes such as LHLH (rise-fall-rise) or HLHL (fall-rise-fall).

Levels are categorised as being high or low according to where they occur in the speaker's local pitch range. The noticeable difference in pitch height of various level tunes within a speaker's own repertoire motivate the distinction between H and L. Although some of the level tunes might be levelled realisations of underlying pitch movement (see results for more discussion, e.g. Section 6.4.2), levels clearly comprise a category of their own.

It is often difficult to assign a level tune as being high or low in a speaker's pitch range if that utterance is brief and the speaker's adjacent utterances are also brief. Additionally, if the pitch range is reduced, a high may not be very different from a low. Distinctions are made on the basis of a best guess, with reference to nearby utterances and the speaker's overall pitch range. It may be the case that some highs will be lower than some lows or vice-versa.

Examples of the five contour types can be seen in Figure 5.1. Difficult classifications are discussed below (Section 5.3.3).

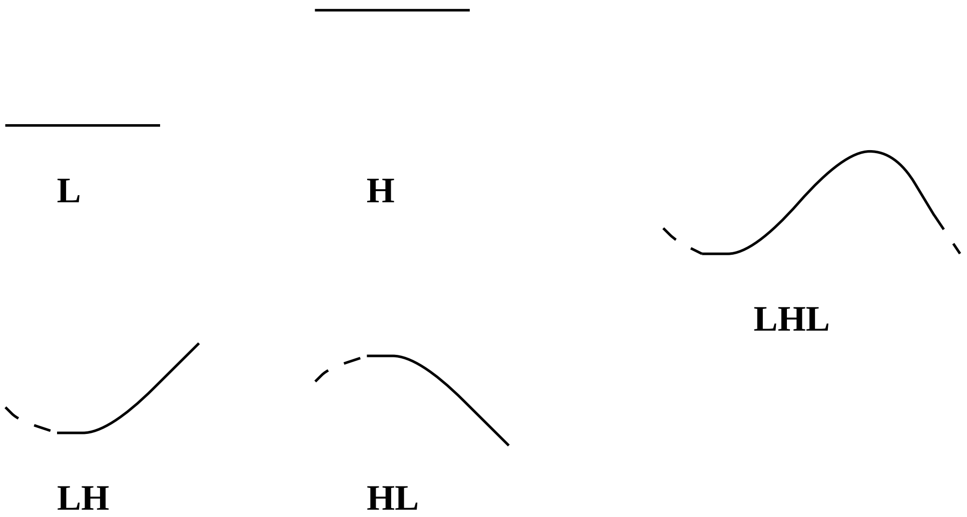

Figure 5.1: Examples of the five contour types

5.3.2 Transcription Procedure

In transcribing, the analyst performs two rounds which effectively make four passes through the data. Primary importance is placed on auditory analysis:

Step 1 Transcribing the data purely through auditory perception

Step 2 ... correcting the transcription by using instrumental analysis

Step 3 Checking the transcription through auditory perception

Step 4 ... checking also with instrumental analysis

The reason for so much checking is that perception of pitch movement can be affected by various artefacts of the utterance. For example, the utterance "yes" which falls significantly in pitch may sound rising because of the high F0 near the fricative [s]. See Figure 5.2.

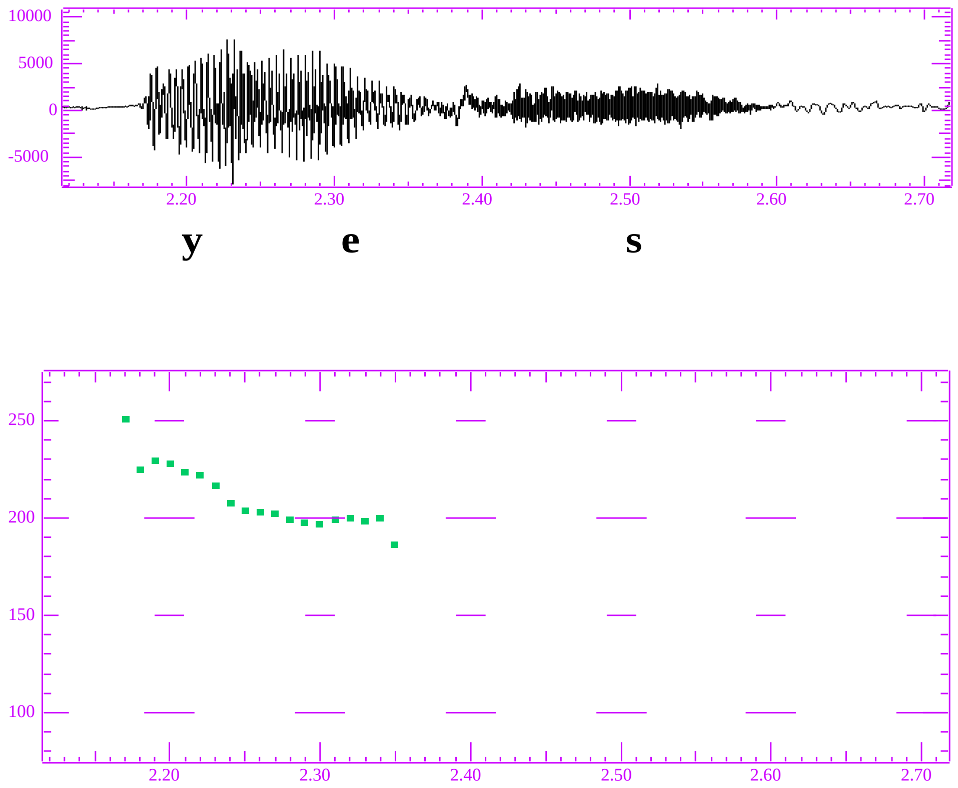

Figure 5.2: "Yes" in which frication masks the fall (top is waveform with Amplitude by Time in seconds, bottom is pitch trace with F0 in Hz by Time)

Auditory Perception (Step 1)

The first step concerns basic labelling: location of accented syllable (usually trivial in the case of single words) and type of accent. Accents are identified by the pitch movement that occurs on and after the perceptually most prominent syllable. An accent that sounds level overall is classified as a level. Although most level accents drift either up or down in pitch (by a few Hz), the drift is present as an artefact of the phonetic utterance and is not part of the accent. (See the section on instrumental analysis, below.) The distinction between H and L is determined by comparison to the pitch range of that same speaker's previous and subsequent utterances. Since a speaker's pitch range may vary from one part of a conversation to another, the pitch range for purposes of H/L distinction is taken from nearby utterances, making it the local pitch range[5].

Identification of the other three contours, HL, LH, and LHL is fairly straightforward. The complex contour LHL is only assigned when three clear targets appear. Some HL accents end in a small dribble of pitch upward but are clearly not identifiable as HLH because the dribble is not significant as a target. The perceived accent is HL.

Instrumental Analysis (Step 2)

Step two of the passes through the data involves analysis of the acoustic signal. It makes reference to digitised versions of the single-word utterances, specifically the waveform (amplitude) and pitch trace. Instrumental analysis involves

[5] As an example of variation in local pitch range, consider two speakers from Quad 5 in the Map Task Corpus. Given three sample regions from a dialogue, Sarah's local pitch range varies from 97 Hz to 179 Hz in width. Danielle's varies from 42 Hz to 73 Hz. Their ranges exhibit the following parameters in the three regions:

Sarah	*High*	*Low*	*Range*	Danielle	*High*	*Low*	*Range*
	280	155	125		138	96	42
	335	156	179		165	92	73
	210	113	97		139	91	48

Sarah's pitch range is always higher and extends further than Danielle's.

checking perception against F0 in terms of

1. Presence of pitch movement

2. Direction of pitch movement

3. Location of H or L target

F0 numerical levels (in Hz) are identified alongside H and L targets. The F0 point is found by aligning the waveform and pitch trace and using techniques inspired by Connell and Ladd (1990: appendix). Peaks and valleys (starting or ending points of falls or rises) are chosen with the following preferences:

1. Probability of voicing = 100%, or where voicing never reaches 100%, 90%+ or the highest number is used. (Probablility of voicing is computed by WAVES software and appears a window alongside the pitch trace. See Section 6.3.2 for details about the software used.)

2. Spurious points including erratic onset and offset points are avoided.

3. Octave errors are corrected.

4. Energy peak is preferred.

Figure 5.3 shows an example of a pitch trace with spurious onset and offset points. Note the level of the probability of voicing (P(voice)) and level of energy (rms).

When an utterance is transcribed as having pitch movement, it tends to change by at least 9 Hz. Levels tend to vary in pitch no more than 8 or 9 Hz either up or down.

Tunes which sound ambiguous between categories and have moderate pitch movement are categorised according to a best guess as to which category they match. In the most ambiguous cases, auditory analysis overrides strict instrumental analysis.

Levels that occur at approximately the middle of a person's pitch range are generally transcribed as low.

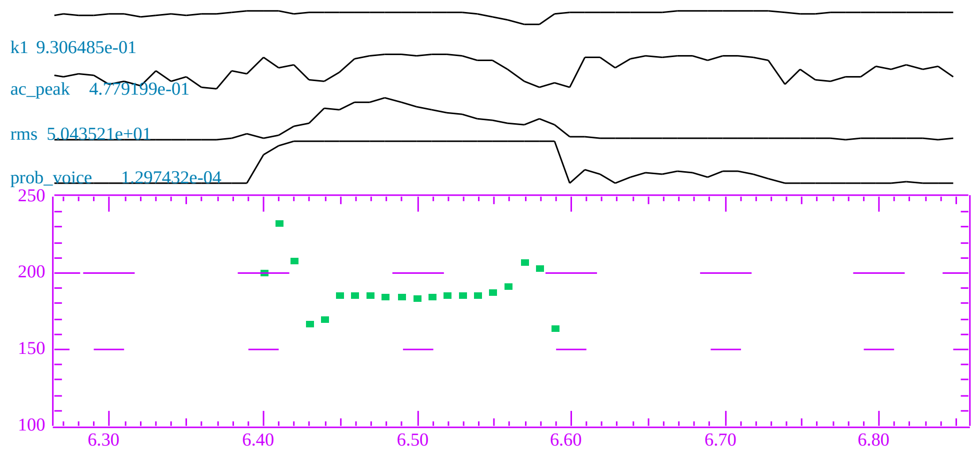

Figure 5.3: Pitch trace of "mmhmm" (low level tune) showing spurious onset and offset points and including probability of voicing and energy levels (F0 in Hz by Time in seconds)

While developing the system, some mid-level target points were identified in the data. These usually appeared in the sequence LH(M). Later it was decided that these were integral to the Glaswegian rise. The M tones were not suffiently low to qualify as L. The tune LH(M) thus became simply LH, along with the simple rises. In Glaswegian, LH(M) often occurs when the final accent covers two equally stressed words or syllables or a lesser stressed word, or words, follow the accent. An unstressed word or syllable may actually carry the H target in LH.

The tunes LH and HL can be characterised as follows:

- LH is usually a simple rising contour, which may have a very short dropped tail. The accented syllable begins either slightly before or during the actual rise, usually at the beginning portion of the rise. There may be a small drop before the rise. Secondary word stress may occur at the peak or on the dropped tail. A variation involves the tail dropping to a point above or almost level to where the rise began.

- HL usually rises slightly at the beginning and ends lower than it began. The accented syllable occurs late in the phrase.

5.3.3 Sorting out some Difficulties

Most utterances are easy to categorise via auditory and instrumental analysis. More than a few utterances are *not* easy to classify. The presence of these is what motivates some of the detail in phonological representation (e.g. high versus low level). Consider two problem utterances which exemplify conflict between auditory and instrumental analyses.

In Figure 5.2, "yes", sounds level and looks falling in pitch. The fricative [s] masks the falling vowel. This utterance is classified as HL because it clearly falls in pitch.

In Figure 5.4, "mmhmm", sounds as if it could be a rise, fall, or level. The contour looks like a moustache. Two main pitch points are considered for the contour shape: the centres of the sloping halves. The moustache shape is noted. The utterance is classified as a high level.

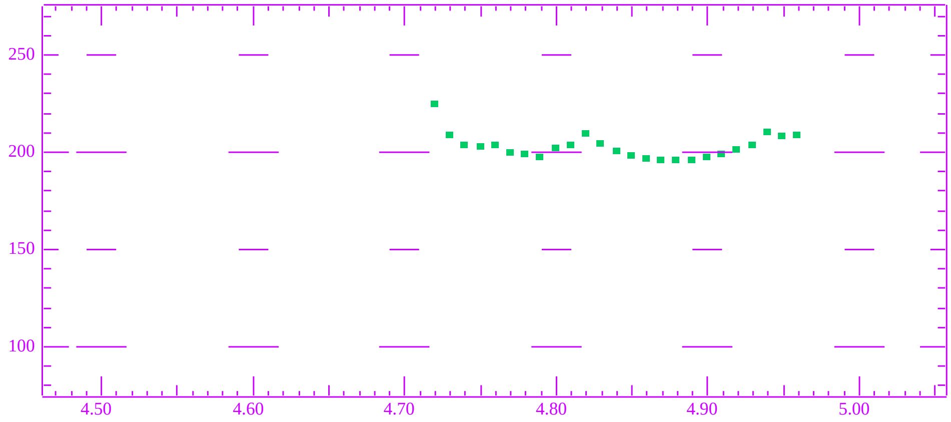

Figure 5.4: "Mmhmm" showing ambiguity (F0 in Hz by Time in seconds)

In the cases of these and other utterances which are difficult to classify, comments are written in the data files with regard to unusual pitch movement.

These comments are considered in the results of the studies in Chapters 6 and 7. More on this subject is discussed in Section 6.3.3.

Chapter 6

Intonation in Discourse: Spontaneous Dialogue

This chapter brings together the two analyses presented earlier as components in a study of how intonation functions within dialogue. It is generally believed that intonation reflects meaning. Although researchers define “meaning” in different ways[1], all agree that intonation encapsulates some linguistic information of its own.

The present study tests the hypothesis shared by other discourse researchers (e.g. Hockey, 1991, 1992; Litman and Hirschberg, 1990; McLemore, 1991) that intonational meaning in some way corresponds to discourse function, and that function is a significant factor in the choice of intonational tune. Although this hypothesis is believed to be true by many contemporary researchers in the field, few claims exist regarding specific intonation behaviour over entire conversations, and the ones that do exist are often sketchy. One difficulty in approaching the problem is deciding how to specify discourse function and the classes of intonational tune.

In the present study, two independent systems, the Conversational Games

[1] For instance, old school language teachers referred to grammatical mood and emotion (e.g. Kingdon, 1958), discourse researchers refer to discourse flow, (e.g. McLemore, 1991), and intonational phonologists identify grammatical, semantic and other distinctions (e.g. Ladd, 1980).

Analysis (Chapter 4) and the description of intonation contours (Chapter 5), form the complementary set of tools for a balanced analysis of dialogue. The Games Analysis provides a framework which represents the function of utterances in dialogue in terms of speaker intentions. The intonation description allows simple but comprehensive transcription of intonation contours in single-word utterances. By using analyses which have been developed independently of one another, we avoid problems encountered in previous work which leaves one analysis less well-defined or dependent upon the other (e.g. Hirschberg and Litman, 1987, 1990; Hockey, 1991, 1992; and McLemore, 1991. See Section 3.4 for a detailed review of these studies). The present work treats both analyses as equally important.

The intonation study below seeks to understand the link between intonation and discourse function in single-word single-function utterances taken from task-oriented dialogue. It addresses two issues:

1. whether discourse function is a significant factor in a speaker's choice of intonation contour in dialogue, and
2. whether intonation strategies are the same in spontaneous and read-aloud dialogue

These issues are covered in separate chapters—the first in the present chapter and the second in Chapter 7. As the materials are the same for both studies, except that one uses spontaneous speech and the other read-aloud speech, a full description of materials will appear only in the present chapter.

6.1 Introduction to the Intonation Study

The intonation study in this chapter looks at single-word conversational moves taken from spontaneous dialogues to determine the relationship between discourse function and type of intonation contour. It tests the following hypothesis:

> Discourse function will correlate significantly with category of intonation contour.

Limiting the data to single-word utterances reduces intonation contours to a manageable size and facilitates an easy comparison between utterances.

Discourse function is defined in terms of the analysis of conversational games and moves (from Chapter 4). The function of an utterance can be specified at different levels by including various amounts of context around the smallest functional unit, the conversational move. For instance, identifying the move without a context provides a minimal description of discourse function (e.g. ACKNOWLEDGE acknowledges something). Identifying the move and the game in which it occurs provides greater specificity about how the utterance functions (e.g. ACKNOWLEDGE in *Querying-YN* acknowledges something to do with a yes-no query). Naming the previous move increases the level of discourse specification again (e.g. ACKNOWLEDGE following QUERY-YN in a *Querying-YN* game indicates that the yes-no question is being acknowledged, not the answer), and so on. The greater the context identified, the more information provided to describe how the speaker intends their utterance to work as a part of the dialogue.

Intonation is represented as a simple contour involving high and low target points (Chapter 5).

Given that the independent representations for discourse function and intonation are adequate, the hypothesis can be rephrased as follows:

> A comparison of the intonation contours of utterances to their discourse function, defined within the framework of the Conversational Games Analysis, will reveal significant correlations.

If discourse function is a significant factor in the choice of intonation contour, then clear patterns of correlation should emerge between the two at some (hopefully low) level of specification. Lack of such patterns would suggest that there are other factors besides discourse context which strongly affect intonation.

This hypothesis makes no prediction regarding the specific intonation contour used by the speaker. It is believed that specific patterns which emerge will depend upon the particular accent studied, but the fact that such patterns appear will be universal. The point of the study is to show that correlations exist between move type and intonation contour, not to guess the nature of the correlations. Other studies have made claims about sentence type and contour (e.g. that questions rise and statements fall), but these generalisations tend to mislead, as they are not very specific about actual discourse context and at best, less specific than the present study. (cf. the literature discussion in Section 3.4.)

6.2 Materials

Materials for the study consist of single-word conversational moves in seven dialogues from the HCRC Map Task Corpus (See Section 4.2 for a description of the Corpus and Section 4.2.8 for detail on the read-aloud dialogues.). The present chapter involves spontaneous dialogue, and the parallel study in the next chapter looks at read speech which arises from the original participants in the Corpus having read aloud from detailed transcripts of their spontaneous dialogues. Although the speech in these dialogues is a variety of Scottish English (Corpus participants speak with a standard Glaswegian accent), any accent of English could be used to test the hypothesis (using an appropriate accent-specific intonation analysis in place of the Glaswegian-specific one in Chapter 5).

6.2.1 Selection

The choice of spontaneous dialogues was constrained by the selection of (matching) read-aloud dialogues available for the parallel study. See Section 4.2.8 for the motivation behind the choice of particular dialogues.

Dialogues were selected from the "no eye contact" quads in the Map Task Corpus to prevent information being communicated via eye gaze. Dialogues without eye contact carry more information through the verbal channel than dialogues with eye contact. Participants could not communicate through body language in the dialogues without eye contact. (This had also been part of the reason that only the "no eye contact" dialogues were re-recorded.)

The final set of dialogues involves participants who had not met previous to the map task. This lack of familiarity avoids the problem of friends interrupting the task with off-topic linguistic exchanges such as jokes.

Male and female speakers are included. The balance is skewed toward females because it reflects the gender balance of the read-aloud quads. Only two male subjects participated in these quads.

The dialogues chosen for the study best meet three additional criteria:

- Good quality speech in the spontaneous and read-aloud renditions
- Correct reading of the transcripts in the read-aloud dialogues
- A natural, spontaneous sound to the read speech

The third criterion, that of a natural sound to the read speech, is important for the comparison of intonation strategies in spontaneous and read-aloud speech. If the read dialogue sounds convincing, then one can say that the speaker realistically interprets utterances, and the comparison between intonation contours in spontaneous and read versions is a reasonable task. Indeed, the read speech of the chosen dialogues is so convincing that upon hearing excerpts from both conditions of dialogue, the experimenter as well as seminar audiences had difficulty in distinguishing which rendition was spontaneous and which was read.

The seven dialogues used, NAQ1C6, NAQ3C1, NAQ3C2, NAQ3C5, NAQ3C6, NAQ5C3, and NAQ5C8, comprise 29 minutes of spontaneous dialogue spoken by 6 female and 2 male speakers. They involve 696 conversational turns (counted in the spontaneous versions).

6.2.2 The Data

All single-word moves from the dialogues were included as candidate data points for the study. Only qualifying data, however, were included in the detailed analysis and ultimately in the results. The following criteria were used to cull data:

- If the word is not in its own conversational turn then it must be in its own intonational phrase.
- The lexical item must be the same in the spontaneous and read versions of the dialogue.
- Speech quality must be good.
- Pitch trace quality must be good.

The data analysed from the seven dialogues include single words which form entire conversational moves. In most cases the move comprises a whole conversational turn. When the move forms part of a longer turn, it always forms its own intonational phrase. Instances such as the following were included in the study because the move is in its own intonational phrase (vertical line here indicates separate intonational phrases; italics denote the single word move):

> *Okay.* | Shall we begin then?
> READY

and

> You're towards the outside of the page? | *Yeah.* | And, | so
> if you've gone round it ... | don don't go beneath it though,
> REPLY-Y

from dialogues NAQ3C1 and NAQ1C6 respectively. The single word moves in these two cases are isolated from the surrounding utterances by clear intonational phrase breaks. Instances of partial intonational phrases were marked as such and not considered in the results. For example, in

Okay, the start's at the top left.

READY

the READY move ("Okay") forms part of the intonational phrase that continues with the EXPLAIN move ("the start's at the top left"). There is no intonational phrase break between the two moves. If a boundary was ambiguous, the relevant data point was excluded from the study. (This example is taken from dialogue NAQ1C6. For context, see the excerpt on page 50 or the whole dialogue in Appendix A).

The final data set is thus comprised of single word moves which form entire intonational phrases, if not entire conversational turns.

Of the 329 single-word moves in the seven dialogues, 29 failed to qualify after examination of the spontaneous dialogues (25 partial intonational phrases, 2 whispered utterances, and 2 instances of laughing obscuring the pitch trace). After examination of the read dialogues, 27 additional utterances were disqualified (13 instances of text differing, 7 partial intonational phrases, 2 barely audible utterances, 2 instances of laughing, 1 wavering utterance, and 2 bad quality pitch traces). The section below gives further detail about how qualifying utterances were identified.

The resulting body of data involves 273 single-word moves (i.e. 83% of all single-word moves) in spontaneous dialogue and a matching set in read-aloud dialogue.

6.3 Method

After the dialogues were selected, a computer file containing orthographic transcription was created for each dialogue. Conversational move and game coding was then added to each file (similar to the sample coded dialogue in Appendix A). Then single words which comprise single moves were marked in the transcripts. These utterances were digitised into waveform files so that data analysis and culling could follow.

New files listing the single-word moves were created and edited to allow the inclusion of information relevant to the hypothesis.

6.3.1 Information Collected

In each dialogue file, the following information can be found for each single-word move. The list of information for each word forms one data record. Note that items 6 and 7 were added *after* the intonation analysis was complete, so that contour classification would not be biased by knowledge of the discourse context. (Appendix B shows sample exerpts of the raw data.)

1. Identity of the turn in terms of
   (a) Dialogue
   (b) Turn number
   (c) Speaker
2. Orthography of the whole conversational turn
3. Target labels (e.g. H, L) and notes concerning ambiguities or phonetic aspects of contour shape
4. F0 levels for each target (frequency in Hertz)
5. Whether the move forms part of a larger intonational phrase (in which case that data point was discarded)
6. Move label (e.g. INSTRUCT)
7. Dialogue context:
   (a) Game in which the move occurs
   (b) Whether the game is embedded
   (c) Move previous to the current one
   (d) Whether the previous move is a continued move
   (e) Subsequent move if spoken by same person in same turn

For purposes of completeness, the study considers all of the above when seeking possible significant interactions with intonation contour, e.g. speaker identity, lexical item, whether or not the previous move is continued from a prior conversational turn, and whether the speaker continues talking in the same conversational turn. Also, for data from conversations for which video recordings exist, the state of the speaker's activity was considered. That is, an active state indicates that the speaker is actively moving across the map, intently looking at it or writing on it, and a non-active state indicates that the speaker is merely talking and thinking.

6.3.2 Computer Processing

For the collection of information in points 3, 4, 5, and 6 above, three files were created for each excerpt of dialogue uttered by a speaker—waveform, pitch trace, and orthography.

Waveform

Speech from each dialogue was digitised from the original recording on digital audio tape (DAT), via analogue output, into 16-bit files readable by a Sun workstation. An Ariel s32c DSP card performed the digital signal processing via a Proport Audio/Digital box, using the ESPS/WAVES software package (versions 4.1 and 2.1). The DAT tapes record dialogue participants on separate left and right channels. One channel at a time, i.e. the speech of one speaker, was digitised into 20-second waveform files.

These waveform files were then edited using WAVES software in a Sun X-windows environment to remove lengths of silence longer than approximately 0.3 seconds. Non-speech sounds, e.g. breathing and pops caused by misplacement and movement of the microphone, were also edited from the file.

Pitch Extraction

Pitch was extracted using the software available in the WAVES package via the *formant* command. Applied to each waveform file, it produced pitch trace files with five tiers of analysis. Three of those tiers were used in the analysis of data:

1. **F0**, estimate of fundamental frequency
2. **prob-voice**, probability of voicing
3. **rms**, rms (root mean square computation of energy) in rectangular window

Pitch extraction predictably erred at points of glottalisation and low probable voicing. Such error did not significantly affect the pitch reading on the voiced segments of the words and thus did not adversely affect the results. Pitch points off by an octave were obvious and easily corrected.

Using the pitch trace, each single-word move was intonationally transcribed as having one of five different pitch accents as described in Section 5.3.2. The LHL (rise-fall) accent does not appear in the data of this study. Four tunes appear: H (high level), L (low level), HL (fall), and LH (rise). Because Glaswegian rises characteristically include a slight fall on the tail, the rises are sometimes transcribed as LH(M).

The pitch trace was sufficiently accurate and reliable to allow straightforward intonation transcription as is described in Section 5.3, with problems being solved (e.g. transcription problems described in Section 5.2.2) as they arose. Some of the more difficult roblems of categorisation are addressed below.

Orthography

Orthographic transcriptions were added using the *xlabel* facility. They contained words roughly chunked into syntactic phrases.

Windows containing waveform, pitch trace and orthography could be displayed simultaneously in alignment.

6.3.3 Analysis of the Data

Intonation transcription of relevant words and phrases was performed by hand, in the first instance by auditory perception (displaying the waveform only) and then by consulting the pitch trace. In general, the threshold which determined that a contour involved pitch change was a shift of 10Hz in fundamental frequency. Spurious F0 points were ignored, such as erratic onset and offset and points of low voicing. Some problems were encountered in transcribing the data (in addition to the ones detailed in Section 5.2.2).

Solving Problems

Ambiguities between the presence and lack of pitch movement were resolved by choosing the analysis preferred by auditory analysis ('by ear') and making a note that there was ambiguity.[2] Many intonation contours which seemed to defy categorisation were labeled as belonging to the nearest possible class and notes were included regarding their visible shape. It was hoped that these notes would capture some intonational properties at a phonetic level which might have been missed by the phonological distinctions mentioned above.[3]

The level tones are particularly interesting. As an example, "Right" often sounds level but has a pitch trace showing a near-perfect smiley-shaped curve, falling then rising. Figure 6.1 shows one such contour. In the trace, the falling and rising portions occur on voiced segments, at points of high probability of voicing, P(voice). The contour was labeled level (L) with an additional "smiley" comment reflecting the shape. "Uh-huh" often took the shape of a double-curved moustache and was labelled thus as well. Figure 6.2 shows a level "uh-huh" which received a "moustache" label. The interesting portions of this contour are the initial fall and final rise, similar to the edges of the smiley shape. Glottalisation or aspiration accounts for the higher pitch points in the centre of the word. The moustache shape contrasts with other shapes such as *v*-shaped rises. Level "uh-huh" sometimes has one or two clearly sloping

[2] Auditory analysis takes priority because conversational participants rely on auditory perception, not instrumental analysis. Instrumental analysis serves as a check, mainly in terms of direction of pitch movement, though sometimes in degree of pitch change. See Section 5.2.2 about auditory versus instrumental analysis.

[3] This was a necessary measure since at the time the intonation analysis was developed, Glaswegian intonation was not entirely understood. A description of Glaswegian intonation has still not been published in a comprehensive form.

syllables. Figure 6.3 shows an instance of "uh-huh" with two sloping syllables (v-shaped). This tune was transcribed as LH.

It is possible that some of these phonetic shapes embody differences in boundary tones. However, given the brevity of such utterances, it becomes difficult to assess what is boundary tone and what is integral pitch accent. These phonetic description labels are included to distinguish the unusual shapes which were classified alongside contours more similar to category prototypes (e.g. simple rise).

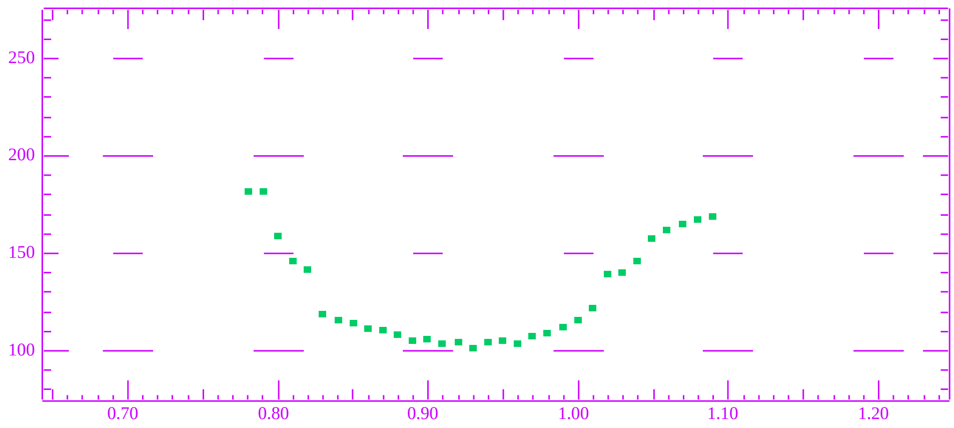

Figure 6.1: Smiley-shaped Level Tone: "Right" (F0 in Hz by Time in seconds)

Collating Data

Data records were collated into groups of particular moves located at particular positions in particular games, and a search for patterns commenced. Since most moves in the data immediately follow an opening move of a game, much of the discourse context can be described as simply move x following move y. All data was compiled by hand and double-checked using computer editors to rearrange the information into a readable form. (See Appendix B for a sample of the raw data.)

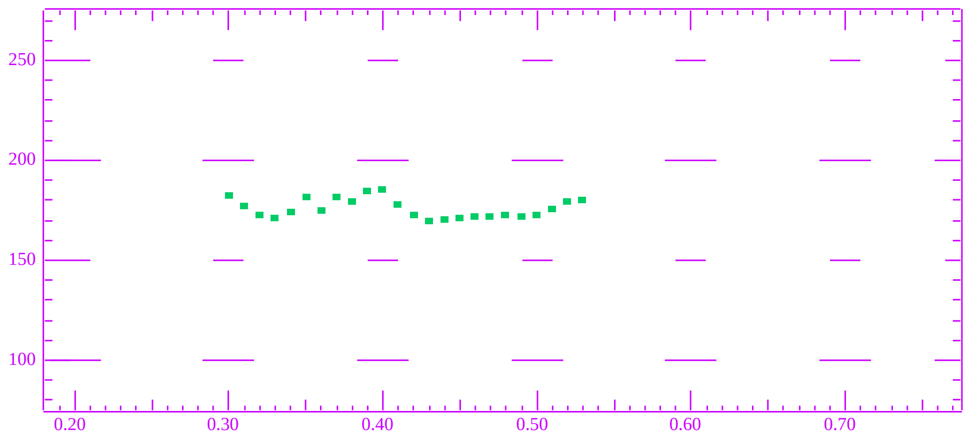

Figure 6.2: Moustache-shaped Level Tone: “Mmhmm” (F0 in Hz by Time in seconds)

6.3.4 Final Data Set

The final data set involves 273 single-word moves which comprise whole intonational phrases. Of the 273 data points, 16% (45) are spoken by male speakers and 84% (228) by female speakers. (Recall that dialogue selection forced a skewing of gender toward female speakers.)

The breakdown of gender and speaker role is as follows. Information givers utter 27% (75) of the data points, 32% (24) of which are from male and 68%

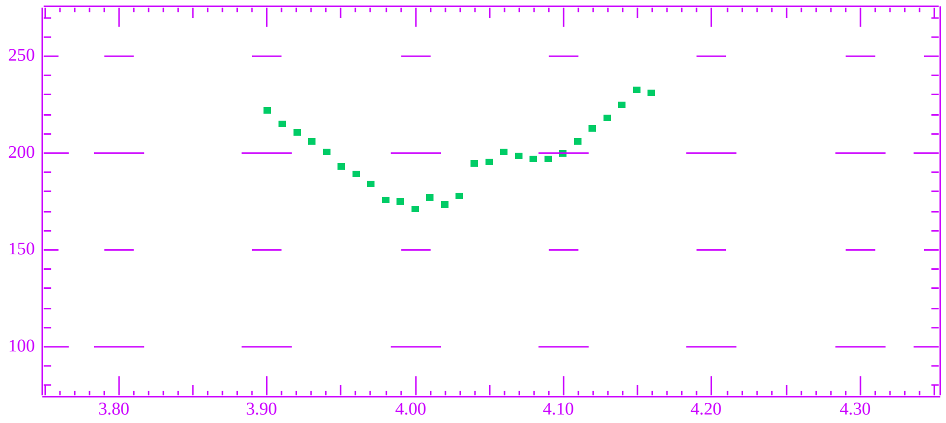

Figure 6.3: V-shaped Rising Tone: “Mmhmm” (F0 in Hz by Time in seconds)

(51) from female speakers. Male speakers are thus overrepresented as givers of information considering their overall contribution to the data. Followers utter 73% (198) of the data points. Male speakers tend to be underrepresented as followers since they utter only 11% (21) as opposed to 89% (177) by females.

The data points comprise 9 of the 12 possible categories of conversational moves: ALIGN, READY, REPLY-Y, REPLY-N, REPLY-W, ACKNOWLEDGE, CHECK, QUERY-W, and QUERY-YN. Table 6.1 shows what number and percentage of data points associate with each move.

Table 6.1: Number and Percentage of the 273 Data Points which Associate with each Move Category ("-" indicates negligible)

Move	#	%
ALIGN	24	9
READY	14	5
REPLY-Y	61	22
REPLY-N	10	4
REPLY-W	1	-
ACKNOWLEDGE	160	59
CHECK	1	-
QUERY-W	1	-
QUERY-YN	1	-

The data set includes 22 words: *ah*, *almost*, *aye*, *cavalry*, *ehm*, *mm*, *mmhmm*, *no*, *nope*, *now*, *okay*, *okey-doke*, *right*, *rightee-ho*, *there*, *uh-huh*, *uh-oh*, *underneath*, *upwards*, *yeah*, *yes*, and *yup*. Table 6.2 shows the number and percentage of data points represented by each lexical item.

Table 6.2: Number and Percentage of the 273 Data Points which Associate with each Lexical Item ("-" indicates negligible)

Word	#	%
Ah	1	-
Almost	1	-
Aye	11	4
Cavalry	1	-
Ehm	1	-
Mm	1	-
Mmhmm	32	12
No	9	3
Nope	2	1
Now	1	-
Okay	45	16
Okey-doke	1	-
Right	97	36
Rightee-ho	1	-
There	1	-
Uh-huh	29	11
Uh-oh	1	-
Underneath	1	-
Upwards	1	-
Yeah	22	8
Yes	12	4
Yup	2	1

6.4 Results

The 273 data points were initially categorised at the lowest functional level, as move x. Intonation contour was compared with discourse context by simply examining the relationship between contour and conversational move. In cases where contour patterns did not appear clearly, it was necessary to raise the level of functional specification and expand the discourse context. The next step was move x in game y. In cases where intonation patterns did not clearly emerge from within the fullest specified discourse context, other possible factors were considered, such as speaker strategies, lexical items, interruptions, and activity level. However these items were expected to (and do) have little or no correlation with intonational contour and were considered only as a last resort in explaining the results. Recall that the hypothesis states that intonation contour choice will significantly correlate with discourse function.

Four intonational categories appeared in the results: rising, high level, low level, and falling. The contours were also collapsed into two categories, "ends high" and "ends low", and other variations of the original four categories which will be mentioned in different subsections where significant.

Statistical significance was tested by the Kolmogorov-Smirnov One-Sample Test (Siegel and Castellan, 1988). This test works on categories with small numbers of data points. Because the process of collecting data for this study takes a long time relative to the usable number of data points produced and time resource was limited, the categories in the study have relatively small numbers of data points. The numbers are too small to undergo a chi square (χ^2) test or an analysis of variance (ANOVA). The Kolmogorov-Smirnov One-Sample Test, however, is designed to produce results based on such numbers of data points.

A probability level of $p = .05$ is the borderline for a statistic being reliably significant. The measure "p" indicates the probability that the pattern tested is due to random behaviour. Lower probabilities thus indicate higher likelihood that the hypothesis is true.

For the statistical tests, data was grouped in different sets of intonation categories. In the first instance, the test checked randomness across all (four) intonation categories. In later instances, the data was grouped to separate out different intonation categories. There was some evidence (discussed later) that LH (Rise) and H (High Level) had a possible link in terms of some H's being underlying LH's. Likewise, some L's may be underlying HL's, and thus H+LH were grouped and L+HL were grouped. Levels (H and L) were also grouped. Significantly non-random distributions were reported at the level of the least number of categories grouped, e.g. L+HL (tunes ending low in pitch) was preferred to H+L+HL (non-rising tunes).

6.4.1 Overall Results

Table 6.3 shows the overall results comparing intonation contour with conversational move.[4] As predicted, patterns emerge from the table showing correlation between conversational moves (embodying utterance function) and intonation contour. Four of the five categories of move which have more than one data point contain a non-random pattern across the four intonational tune categories. The non-random pattern also appears in the total for all moves.

Most of the tunes which appear in the results involve pitch movement (65%, trend at $p < .10$): one-third (33%) rise and one-third (32%) fall in pitch. Level tunes account for the other third of the total (14% high and 21% low). Complex tunes, rise-falls, do not appear in the data from this study.

6.4.2 Align Moves

Utterances classified as align moves correlate significantly with rising intonation contours (79%, $p < .05$). It is possible that the five cases of level contour may be ones in which the underlying tune would be a rise but was never realised as such. Three of the level tunes provide some evidence for adopting this view.

[4] Tunes are presented and discussed here in terms of pitch movement because "rise" and "fall" are visually easier to interpret than a succession of L's and H's in the thesis text. Use of such terms is not meant to detract from the representation of tune in terms of target levels.

Table 6.3: Spontaneous Dialogue Results: Intonation Contour Associated with Conversational Move (‡ indicates significantly non-random distribution across four categories at $p < .01$, † at $p < .05$)

		Levels			
Move	*Rise*	*Hi*	*Lo*	*Fall*	*Total*
ALIGN‡	19	3	2		24
READY	2	2	5	5	14
ACKNOWLEDGE‡	63	23	35	39	160
REPLY-Y‡	4	7	14	36	61
REPLY-N†		1	2	7	10
REPLY-W		1			1
CHECK				1	1
QUERY-YN	1				1
QUERY-W	1				1
Total	90	37	58	88	273

In two of the instances, the fundamental frequency rises by 4 Hz. One of these occurs immediately after a monotone phrase, adding credence to the idea that 4 Hz marks an intended, significant pitch change. A third instance involves a high level tone which immediately follows an intonational phrase ending at a (10 Hz) lower fundamental frequency. The low ending to the phrase immediately previous suggests that the high level may abbreviate the end of a rising contour.

6.4.3 Ready Moves

The READY moves, although not significantly different from random across the four tune categories ($p > .20$), are significantly non-rising (86%, $p < .05$) when tunes are grouped as two categories, rising and non-rising. The two rising tunes are secondarily labelled as having a "smiley" shape rather than a strict rise (recall the discussion of contour shape in Section 6.3.3 above) and include comments that the contours sound ambiguous and may be classified as low level.

Five of the six "smiley" or ambiguous tunes (including the two rises) in these results occur after an ACKNOWLEDGE move. This context may explain the presence of the rising tunes. See Table 6.4 for the correlation between intonation and discourse context which includes previous move.

Speaker strategy might also explain the pattern of tunes, as one person utters six levels, two rises and one fall. The other five utterances (four falls and a high level) are spoken by three individuals.

Numbers are too small to decide between any of the above explanations. Regardless, it is clear that the significant majority of tunes are non-rising.

Table 6.4: Spontaneous Dialogue Results: Ready Moves (Statistics hold no significance at $p > .15$)

Move	*Prev. Move*	*Rise*	*Hi*	*Lo*	*Fall*	*Total*
READY	ACKNOWLEDGE	2		3	1	6
	REPLY-Y		1	1		2
	ALIGN		1			1
	EXPLAIN				1	1
	–			1	3	4

6.4.4 ACKNOWLEDGE **Moves**

The intonation pattern for ACKNOWLEDGE moves shows 39% rises, 36% levels (14% high and 22% low), and 24% falls, a non-random spread across the four categories of pitch movement ($p < .01$). Most moves are non-falling (76%, significant at $p < .01$). A slightly clearer picture emerges by looking at a higher specification level of discourse context. It is worth noting at this point that of the 21 ACKNOWLEDGE moves (in different conversational games) after which the same speaker continues talking in their conversational turn, a significant majority end low in pitch (81%, $p < .05$; 2 rising, 2 high level, 10 low level, and 7 falling are non-random at $p < .05$). A speaker generally utters tunes that end low before continuing to talk in that conversational turn.

Table 6.5: Spontaneous Dialogue Results: Acknowledge Moves (‡ indicates significantly non-random distribution at $p < .01$)

Move	*Game*	*Rise*	*Hi*	*Lo*	*Fall*	*Total*
ACKNOWLEDGE	*Instructing*‡	51	18	22	20	111
	Explaining	3		3		6
	Aligning				2	2
	Checking	2		3	7	12
	Querying-YN	4	5	5	8	22
	Querying-W	3		2	2	7
ACKNOWLEDGE	embedded *Querying-YN*	4	1	3	5	13
	top-level *Querying-YN*		4	2	3	9
	embedded *Querying-W*	3		2	2	7

Table 6.5 (top part) shows the contours which associate with ACKNOWLEDGE moves in more specific context. From this table it is apparent that the moves in *Instructing* games are significantly non-falling (82%, $p < .01$). There may be some speaker effect, as when the two speakers from Quad 5 dialogues are considered separately from the others, the Quad 5 speakers show no preference for rising and falling tunes (11 rises, 13 falls) while the other six speakers mostly end low in tune (77%, significant at $p < .01$). There is no lexical effect which

explains the presence of the falling tunes. In *Explaining* games the moves are all non-falling (trend at $p < .10$), and in *Aligning* games they fall. ACKNOWLEDGE moves in *Checking* games end low (83%, significant at $p < .05$). In *Querying-YN* games they are non-rising (82%, significant at $p < .05$).

We need to look at whether or not games are embedded to make better sense of the *Querying* game results. The bottom part of Table 6.5 shows the breakdown of *Querying* games into embedded and top-level categories. The rising tunes are only present in embedded games. It is in these embedded games that tunes are spread almost evenly between rising, level, and falling. In the non-embedded, or top-level, *Querying* games (in this case *Querying-YN*), ACKNOWLEDGE moves are non-rising (100%, significant at $p < .05$). There is some indication that in the embedded *Querying* games speaker strategy is accounting for the spread of tunes. Half of the speakers (three) contribute 10 non-falling tunes (100%, significant at $p < .01$) and half (three) contribute 9 non-rising tunes (90%, trend at $p < .10$).

Note that in the *Querying* games, the ACKNOWLEDGE move is almost always spoken by the person who initiates the game (93%, 29 data points; significant at $p < .01$). The following excerpt from dialogue NAQ3C1 illustrates this. Speaker G is the information giver and F the information follower:

G And straight down ehm to ...

INSTRUCT–cont

–begin embedded Querying-YN game–

Do you have a trout farm?

QUERY-YN

F Yes, I've got a trout farm.

REPLY-Y

G Uh-huh.

ACKNOWLEDGE

–end embedded Querying-YN game–

Ehm sort of south down and to horizontally along underneath the trout farm. So you're going down and then

INSTRUCT–cont

This contrasts with moves in *Instructing* games which are almost always *not* spoken by the initiator of the game (97%, 111 data points; significant at $p <$ .01). The following example immediately precedes the above one from dialogue NAQ3C1:

–begin Instructing game–

G Ehm, round above horizontally over above the gold mine.

INSTRUCT

F Okay.

ACKNOWLEDGE

The results can be described in terms of whether or not the move is uttered by the game initiator. Table 6.6 shows that ACKNOWLEDGE moves spoken by the initiator of a game are non-rising (81%, significant at $p < .01$). When spoken by the responder, they are non-falling (82%, significant at $p < .01$). If moves are separated into those which occur in top-level games and those which occur in embedded games, all of the rising tunes spoken by the game initiator appear in embedded games. The same thing cannot be said for the game responder (only one rising tune is found in embedded games). Otherwise, the patterns for embedded games are similar to those for top-level games.

Table 6.6: Spontaneous Dialogue Results: Acknowledge Moves (‡ indicates significantly non-random distribution at $p < .01$)

ACKNOWLEDGE *move*		*Levels*			
spoken by	*Rise*	*Hi*	*Lo*	*Fall*	*Total*
Game Initiator‡	8	6	11	18	43
Game Responder‡	55	17	24	21	117

Table 6.7 shows the discourse context for ACKNOWLEDGE moves further specified to include previous move. This however adds no appreciable understanding to the results.

6.4.5 REPLY-Y Moves

REPLY-Y moves are significantly non-rising ($p < .01$) and indeed end low in pitch ($p < .01$). Almost all low level tunes fall slightly in pitch. Of the four rises, two are labelled as possibly being low level tunes. The two clear rises in pitch (20 Hz or more) have the secondary label "v-shape" because the contour actually falls before rising.

It is perhaps more illuminating to expand the discourse context. Table 6.8 considers the game in which the move occurs. None of the rises occur in *Querying-YN* games (significantly ending low at $p < .01$). They are found in games where the level of expectation is higher that a positive response will be given (i.e. in *Checking* and *Aligning* games).

6.4.6 REPLY-N Moves

The REPLY-N moves are comprised of 70% falling and 30% level tunes (trend at $p < .10$). The intonation pattern mimics that of REPLY-Y moves, i.e. significantly non-rising (100%, $p < .01$; trend at $p < .10$ for ending low in pitch). In fact, the significant majority of all REPLY moves end low in pitch (82%, $p < .01$).

6.4.7 Other Moves

The remaining moves are REPLY-W, CHECK, QUERY-YN, and QUERY-W. Not much can be said for categories with one data point each. Note, however, that the reply is non-rising (as are the other replies) and that both queries rise in tune.

6.4.8 Lexical Item

It may interest the reader to see which lexical items associate with particular intonation contours and particular moves. The hypothesis does not address lexical item because no lexical effect was expected in the results.

Table 6.9 shows that five words have significantly non-random tune patterns in the data. "Aye", "no", and "yeah" have non-rising tunes (significant at $p < .01$, $p < .05$ and $p < .01$ respectively), while "mmhmm" and "okay" have non-falling tunes (both significant at $p < .01$).

Table 6.10 shows that the only words which ever associate with more than two types of move are "okay" and "right". The lexical items which appear in two types of move display a greater than 2:1 preference for one of the moves.

The strong correlation between lexical item and discourse category suggests that there is little difference between the two and the probability is high that we can safely ignore the identity of lexical items in this study. In fact, the results for discourse categories (shown below) individually support this measure and fail to show any lexical effect.

Table 6.7: Spontaneous Dialogue Results: Some Acknowledge Moves in Greater Detail (‡ indicates significantly non-random distribution at $p < .01$)

ACKNOWLEDGE *Moves*			*Levels*			
Game	*Previous Move*	*Rise*	*Hi*	*Lo*	*Fall*	*Total*
Instructing	INSTRUCT‡	48	15	21	19	103
	CLARIFY	2				2
	ACKNOWL.	1	3	1	1	6
Checking	REPLY-Y	1		3		4
	REPLY-W				1	1
	CLARIFY	1			4	5
	EXPLAIN				1	1
	ACKNOWL.				1	1
Querying-YN	QUERY-YN		1			1
	REPLY-Y	2	1	4	3	10
	REPLY-N	1	2		1	4
	REPLY-W				1	1
	CLARIFY	1	1	1	2	5
	ACKNOWL.				1	1
Querying-W	QUERY-W	1				1
	REPLY-W	2		2	2	6

Table 6.8: Spontaneous Dialogue Results: Reply-Y Moves (‡ indicates significantly non-random distribution at $p < .01$, † at $p < .05$)

Move	*Game*	*Rise*	*Hi*	*Lo*	*Fall*	*Total*
REPLY-Y	*Querying-YN*‡		4	7	18	29
	Checking	1	2	3	7	13
	Aligning†	3	1	4	11	19

Table 6.9: Lexical Item and Spontaneous Intonation Contour (‡ indicates significantly non-random distribution at $p < .01$, † at $p < .05$)

		Levels		
Word	*Rise*	*Hi*	*Lo*	*Fall*
Ah				1
Almost		1		
Aye†		1	3	7
Cavalry		1		
Ehm	1			
Mm	1			
Mmhmm†	15	10	2	5
No†		1	2	6
Nope		1		1
Now			1	
Okay‡	29	7	2	7
Okey-doke	1			
Right	33	8	35	21
Rightee-ho	1			
There				1
Uh-huh	6	5	6	12
Uh-oh				1
Underneath	1			
Upwards				1
Yeah‡	1	2	6	13
Yes	1		1	10
Yup				2

Table 6.10: Lexical Item and Conversational Move (align, ready, acknowledge, reply-y, reply-n, reply-w, check, query-yn, and query-w)

Word	*Al*	*Re*	*Ac*	*R-Y*	*R-N*	*R-W*	*Ch*	*Q-YN*	*Q-W*
Ah			1						
Almost						1			
Aye			3	8					
Cavalry			1						
Ehm			1						
Mm									1
Mmhmm			25	7					
No			1		8				
Nope					2				
Now		1							
Okay	16	1	25	3					
Okey-doke			1						
Right	8	12	75	2					
Rightee-ho			1						
There			1						
Uh-huh			20	9					
Uh-oh			1						
Underneath								1	
Upwards							1		
Yeah			2	20					
Yes			1	11					
Yup			1	1					

6.5 Discussion

The results show clearly that discourse function is a significant factor in choice of intonation contour within dialogue. They also suggest that the Games Analysis is an effective system of dialogue analysis. A summary of results can be seen in Table 6.11. ALIGN moves correlate with rising intonation contours, and READY moves with non-rising tunes. ACKNOWLEDGE moves overall are non-falling but only hold this pattern in *Instructing* and *Explaining* games. The pattern in *Aligning Checking*, *Querying-YN* and all top-level *Querying* games is non-rising. (Actually it is falling in *Aligning* games and ends low in *Checking* games.) There is some speaker effect in the embedded *Querying* games. Also, when an ACKNOWLEDGE move is uttered before that speaker has finished his or her conversational turn, it will end low in pitch. When ACKNOWLEDGE moves are spoken by the game initiator, they are non-rising. When spoken by the responder, they are non-falling. REPLY-Y moves end low in pitch. REPLY-N moves are non-rising. (No rises associate with the REPLY-Y moves in *Querying-YN* games.) REPLY moves overall end low.

The resulting tune patterns suggest possible links between certain kinds of moves. For example, READY moves and ACKNOWLEDGE moves spoken before the conversational turn continues occur before the same speaker continues talking. In these moves, the tunes are non-rising, tending toward ending low in pitch. One might call these instances of a continuation situation (e.g. Brown, Currie and Kenworthy, 1980; McLemore, 1991) but here instead of a continuation rise, we have a continuation non-rise or perhaps a continuation 'ends low'.

We can see that the type of game affects contours in ACKNOWLEDGE moves. In information-giving games (e.g. *Instructing* and *Explaining*) the acknowledgements do not often fall in pitch. In information-seeking games they tend to fall in pitch or at least not rise. Acknowledgements in information-seeking games are almost always spoken by the game initiator. In information-giving games they are almost always spoken by the other participant. When considering the

Table 6.11: Summary of Spontaneous Dialogue Results: Discourse Functions which Associate with Tunes ($\star$ indicates pitch ends low)

Tunes			
Rising	*Non-Falling*	*Non-Rising*	*Falling*
ALIGN	ACKNOWLEDGE	READY	ACKN. in *Aligning*
	ACKN. in *Instructing*	ACKN. in *Checking* $\star$	
	ACKN. in *Explaining*	ACKN. in *Querying-YN*	
	ACKN. spoken by game responder	ACKN. in top-level *Querying* games	
		ACKN. before continuing in same conversational turn	
		ACKN. spoken by game initiator	
		REPLY-Y $\star$	
		REPLY-N	
		REPLY $\star$	

results in terms of speaker role (i.e. initiator or responder) as opposed to move context (game in which it occurs), some information is lost, namely that acknowledgements in *Aligning* games fall in tune. Therefore, game context is the preferred analysis rather than speaker role. Interestingly, the REPLY moves, present in information-seeking games and never uttered by the game initiator, tend to fall in pitch.

ALIGN moves are in fact queries relating to the listener's state of attention and understanding. They rise in pitch.

It is possible that some of the variation of tune with ACKNOWLEDGE moves is due to a confusion between ACKNOWLEDGE and READY moves. Both are feedback moves with little content and often serve to ground the other speaker's utterance. When an acknowledgement ends a game it can also function as a transitional READY move.

There is no evidence of two of the possible factors interacting with intonation: whether the previous move has a –cont feature (See Section 4.3.4) and whether the speaker is engaged in activity while the utterance is spoken.

Although researchers have traditionally sought results showing correlation between discourse categories and specific intonation categories, the results in this study, and indeed in other research, show that single intonation categories are not necessarily appropriate. Lisker (p.c.) conducted a study involving native speakers reading aloud non-English texts several times. He noted that the pitch accent types for particular sentences were generally in free variation excepting one category. That is, speakers reading aloud text have rules for intonation assignment which generally involve categories of exclusion rather than single categories. Based upon results in the present chapter, a similar phenomenon occurs for some functional units of spontaneous speech. Possible reasons for this are considered in the Conclusion (Chapter 8).

It remains that discourse function correlates significantly with intonation contour, in accordance with the hypothesis.

Chapter 7

Intonation in Discourse: Read Dialogue

This chapter presents a study of how intonation functions within read-aloud dialogue. It is an extension of the study on spontaneous dialogue in the previous chapter.

7.1 Introduction to the Read Speech Intonation Study

The present study addresses the second issue mentioned in the previous chapter (page 99), namely whether intonation strategies are the same in spontaneous and read-aloud dialogue. Correlation has already been established between units of discourse function and intonation contour in spontaneous dialogue. This chapter performs a similar examination of read-aloud dialogue.

A study was conducted using the read-aloud dialogues identical in terms of text (and participants) to those in the spontaneous study from the previous chapter. Similar methodology was employed, comparing discourse function as specified by the Conversational Games Analysis (Chapter 4) to intonation categories (as in Chapter 5).

The hypothesis in this study is two-fold. The first part is identical to the hypothesis in the previous chapter except that here it applies to speech from read-aloud dialogue:

> A comparison of the intonation contours of utterances to their discourse function, defined within the framework of the Conversational Games Analysis, will reveal significant correlations.

The second part addresses the issue of read versus spontaneous speech:

> Read speech will show slightly better patterns of correlation than will spontaneous speech.

Because read dialogue involves different processes from spontaneous dialogue, it is expected that different patterns of correlation will appear between intonation contour and discourse function. For instance, read-aloud dialogue involves reading, interpreting and uttering conversational turns. In spontaneous dialogue the main task of the speaker is to copy the path on the map, not to read a transcript. The speakers of the read dialogues assign a functional interpretation to utterances in the transcript during their reading aloud. This results in their renditions sounding convincingly like the original spontaneous version. Using their implicit (procedural) knowledge of which intonation contours are appropriate, or typical, for the contexts, they utter the phrases with these appropriate contours. Their rules for contour assignment are expected to approximate a clear mapping, and as such, the results for read speech are expected to show better (more homogenous) patterns of correlation than in spontaneous speech.

Having two sets of data from dialogues which are identical, aside from the fact that one is obtained spontaneously and the other by reading aloud, allows us to assess another piece of information. By examining the level of agreement between individual data points across the dialogue conditions, we can determine whether subjects implicitly agree with the functional categorisation available in the Conversational Games Analysis or whether they are using other strategies which correlate better with intonation contour. The main hypothesis predicts the former, that the subjects agree with the Games Analysis. This comparison study will be presented in Section 7.7 after discussion of the main study.

7.2 Materials

Materials for this study are identical to that in the previous chapter (Section 6.2) with the exception that read-aloud dialogues associated with the HCRC Map Task Corpus are used instead of spontaneous dialogues from the Corpus (See Section 4.2 about the Corpus.). The dialogues have undergone the same selection process as in the spontaneous speech study. In summary, these materials include seven dialogues containing Glasgow Scottish English. They meet the selection criteria described in Section 6.2.1.

The seven dialogues, NAQ1C6, NAQ3C1, NAQ3C2, NAQ3C5, NAQ3C6, NAQ5C3, and NAQ5C8, comprise 24 minutes of read dialogue spoken by 6 female and 2 male speakers. They involve 696 conversational turns (counted in the spontaneous versions).

The data set involves single-word conversational moves which comprise their own intonational phrases.

7.3 Method

The method involved in collecting and analysing relevant data points is also identical to the study in the previous chapter. See Section 6.3 for details.

The final data set therefore involves 273 single-word conversational moves. See Section 6.3.4 for breakdown of gender and speaker role. The same section also lists tables showing the number of data points in each category of conversational move (Table 6.1) and lexical item (Table 6.2). The data points involve the identical 9 of 12 possible conversational moves: ALIGN, READY, REPLY-Y, REPLY-N, REPLY-W, ACKNOWLEDGE, CHECK, QUERY-W, and QUERY-YN.

7.4 Results

Results are presented as in the previous chapter. Each category of discourse function is analysed in turn. The 273 data points were initially classed according to intonation contour and discourse context represented by the conversational move. When necessary, discourse context was expanded to include the game and previous move. In cases where intonation patterns do not fully emerge from within the fullest specified discourse context, other factors are considered, such as speaker strategies, lexical items, and interruptions. Significance of results is reported according to the Kolmogorov-Smirnov One-Sample Test (Siegel and Castellan, 1988) applied to data as described in Section 6.4.

7.4.1 Overall Results

Overall results can be seen in Table 7.1.[1] The largest four categories of move show significantly non-random patterns across all of the intonation categories, as does the total for all moves. The first four categories and the total also show a significantly non-random balance when the intonation categories are collapsed into "ends high" and "ends low" ($p < .01$).

Almost half of the intonation contours in these results fall in pitch (45%). One-quarter (24%) rise while another approximate quarter (29%) remain level. Of the level tunes, the majority are low in the speaker's local pitch range (21% low as opposed to 9% high). Few complex tunes appear (1%). They are the rise-falls characteristic of longer utterances in the Glasgow accent (see Section 3.3).

7.4.2 ALIGN Moves

ALIGN moves show a significant correlation with intonational tune. They almost all rise in pitch (92%, $p < .01$). It is probable that the single falling tune is a case of mistaken interpretation of the utterance as an acknowledgement instead

[1] See footnote on page 114 regarding the lack of "H" and "L" in the tables and text of this chapter.

Table 7.1: Read Dialogue Results: Intonation Contour Associated with Conversational Move (‡ indicates significantly non-random distribution at $p < .01$, † at $p < .05$)

		Levels				
Move	*Rise*	*Hi*	*Lo*	*Fall*	*Rise-Fall*	*Total*
ALIGN‡	22		1	1		24
READY‡			2	12		14
ACKNOWLEDGE‡	36	15	31	77	1	160
REPLY-Y‡	6	7	21	27		61
REPLY-N	1	2	2	5		10
REPLY-W			1			1
CHECK					1	1
QUERY-YN					1	1
QUERY-W				1		1
Total‡	65	24	58	123	3	273

of an aligning query. In any case a comment suggests that the fall may be a high level tune – an upside-down smiley shape. The low is noted as a “flattened glasgow ending”, an underlying rise spoken in either a very narrow pitch range or a rather monotone voice.

7.4.3 Ready Moves

Ready moves significantly fall in tune (86%, $p < .05$).

7.4.4 Acknowledge Moves

Utterances classed as ACKNOWLEDGE moves display a more complex pattern. The majority of tunes end low in pitch (68%, $p < .01$). The moves which occur before a speaker has finished their conversational turn primarily end low in pitch (86%, significant at $p < .01$; 1 rise, 2 high levels, 4 low levels, 13 falls, 1 rise-fall).

By looking at the game in which each utterance occurs, we can see a clearer pattern. Table 7.2 shows the results. ACKNOWLEDGE moves in *Aligning* and *Checking* games exclusively fall in tune (significance in the *Checking* games at $p < .01$). The moves in *Explaining* games are non-rising (trend at $p < .10$. Moves in *Querying* games end low in pitch: 91% in *Querying-YN* games (significant at $p < .01$) and 71% in *Querying-W* games. The rising tunes in *Querying* games occur when the games are embedded. ACKNOWLEDGE moves in top-level *Querying-YN* games end low in pitch (100%, significant at $p < .05$).

Almost all of the rising tunes occur in *Instructing* games. This last category has tunes which almost perfectly split three ways into rising (31%), level (33%) and falling (36%) tunes. Looking at the identity of the previous move adds something to clarify the pattern. Acknowledgements which do not follow instructions are significantly non-rising (100% of 8 data points, significant at $p < .05$). When they follow instructions, ACKNOWLEDGE moves can be described as either non-rising or non-falling (67% and 66%, significant at $p < .01$ and $p < .05$, respectively), i.e. an ambiguous pattern. Whether the ACKNOWLEDGE is following an initial instruction or a continued instruction appears to make no difference – the pattern remains split almost evenly into rises, levels and falls (23 rises, 7 high and 15 low levels, and 20 falls; and 11 rises, 3 high and 9 low levels, and 14 falls, respectively). No further clarity in the results is obtained by looking at discourse context which includes other previous moves.

There appear to be other effects present in the results of the *Instructing* games, specifically some speaker strategy and lexical effect. Three of the eight speakers utter almost no rising tunes in *Instructing* games (94% of 32, non-rising, significant at $p < .01$; can also be considered significantly ending low, 75%, $p < .05$). The lexical item "mmhmm" correlates with non-falling tunes (95% of 22, significant at $p < .01$) and contributes 12 of the 34 rises. "Right" is mostly non-rising (80% of 50, significant at $p < .01$). These are the only two lexical items which exhibit non-random patterns.

Table 7.2: Read Dialogue Results: Acknowledge Moves (‡ indicates significantly non-random distribution at $p < .01$, † at $p < .05$)

Move	*Game*	*Rise*	*Hi*	*Lo*	*Fall*	*Rise-Fall*	*Total*
ACKNOWLEDGE	*Instructing*‡	34	11	26	40		111
	Explaining		2	1	3		6
	Aligning				2		2
	Checking‡				12		12
	Querying-YN‡	1	1	3	16	1	22
	Querying-W	1	1	1	4		7
ACKNOWLEDGE	embedded *Querying-YN*‡	1	1	2	9		13
	top-level *Querying-YN*‡			1	7	1	9
	embedded *Querying-W*	1	1	1	4		7

The results for ACKNOWLEDGE moves can also be described in terms of speaker role. Table 7.3 shows that ACKNOWLEDGE moves spoken by the game initiator have a significant majority of falling tunes (77%, $p < .01$). Those spoken by the responder are significantly non-rising (71%, $p < .01$). The two rises uttered by the game initiator occur in embedded games. No rises occur when moves are uttered by the responder in embedded games. Looking at the results in this manner does not clarify patterns found above. Defining discourse function in terms of the Conversational Games Analysis enables more specific results to appear.

Table 7.3: Read Dialogue Results: Acknowledge Moves (‡ indicates significantly non-random distribution at $p < .01$)

ACKNOWLEDGE *move*		*Levels*				
spoken by	*Rise*	*Hi*	*Lo*	*Fall*	*Rise-Fall*	*Total*
Game Initiator‡	2	2	5	33	1	43
Game Responder‡	34	13	26	44		117

7.4.5 REPLY-Y Moves

The majority of REPLY-Y moves end low in pitch (79%, significant at $p < .01$). It is possible that some of the rises are interpretation mistakes in the reading task. One rise is commented to be uttered "like a question" instead of a reply. The rises cannot be explained by conversational game context, speaker strategy, or lexical effect.

7.4.6 REPLY-N Moves

REPLY-N moves exhibit a non-rising trend (90%, $p < .10$). The single rise "may be high level" and is commented as such.

7.4.7 Other Moves

The other four moves have intonation contours which end low in pitch. Note that all of the REPLY moves end low in pitch (significance at $p < .01$).

7.4.8 Lexical Item

As in the previous chapter, a summary of lexical item and intonation contour is provided here (Table 7.4) for the reader's interest. Although there appears to be some non-random patterning between three of the words ("mmhmm", "okay", and "right") and the categories of conversational tune, the results above show no significant lexical effect aside from possibly "mmhmm" and "right" in ACKNOWLEDGE moves within *Instructing* games. As Table 6.10 in the previous chapter shows, lexical items tend not to associate with more than one category of discourse function.

Table 7.4: Lexical Item and Read Intonation Contour (‡ indicates significantly non-random distribution at $p < .01$)

		Levels			
Word	*Rise*	*Hi*	*Lo*	*Fall*	*Rise-Fall*
Ah				1	
Almost			1		
Aye	1		6	4	
Cavalry					1
Ehm		1			
Mm				1	
Mmhmm‡	13	10	5	4	
No		1	3	5	
Nope	1	1			
Now				1	
Okay‡	23	1	2	19	
Okey-doke				1	
Right‡	17	5	21	54	
Rightee-ho	1				
There				1	
Uh-huh	6	2	7	14	
Uh-oh				1	
Underneath					1
Upwards					1
Yeah	2	2	7	11	
Yes	1	1	5	5	
Yup			1	1	

7.5 Discussion

As with the spontaneous speech, the results for read speech show that discourse function correlates significantly with a speaker's choice of intonation contour. Table 7.5 displays a summary. ALIGN moves almost all rise in tune. READY moves fall, and REPLY moves end low (78%, $p < .01$). ACKNOWLEDGE moves mostly end low in pitch except in *Instructing* games where there is an almost even spread of tunes across the categories of rising, level, and falling. In *Aligning* and *Checking* games they fall, in *Explaning* games they are non-rising, in *Querying* games they end low. The rising tunes in *Querying* games only occur in embedded games. In top-level *Querying* games (actually *Querying-YN*) they end low in pitch. The moves in *Instructing* games which do not follow instructions are significantly non-rising. The *Instructing* game harbours some speaker strategy and lexical effect, for three of the eight speakers and two of the eleven lexical items. ACKNOWLEDGE moves which occur before the speaker has finished the conversational turn end low in pitch. When spoken by the game initiator, the majority of ACKNOWLEDGE moves fall in tune. They are non-rising when spoken by the game responder. REPLY-Y moves end low in pitch. REPLY-N moves are non-rising.

As with the spontaneous results, we can see links between certain cagetories of moves by their tune patterns. The moves which occur before the same speaker continues talking, READY moves and ACKNOWLEDGE moves spoken before the conversational turn continues, end low in pitch.

Acknowledgements in information-giving games (from *Explaining* and *Instructing* games excepting those which follow INSTRUCT moves) are non-rising. In information-seeking games, the ACKNOWLEDGE moves veer more toward falling and low level tunes. This division can be explained in terms of speaker role. Game responders (involving almost all moves in *Explaining* and *Instructing* games) produce non-rising tunes while initiators produce falling tunes. The REPLY moves are also present in information-seeking games, but REPLY-Y moves end low while REPLY-N moves are slightly more spread, in a non-rising pattern.

Table 7.5: Summary of Read Dialogue Results: Discourse Functions which Associate with Tunes

Tunes			
Rising	*Non-Rising*	*Ends Low*	*Falling*
ALIGN	ACKN. in *Explaining*	ACKNOWLEDGE	READY
	REPLY-N	ACKN. in *Querying*	ACKN. in *Aligning*
	ACKN. in *Instructing* which do not follow instructions	ACKN. before continuing in same conversational turn	ACKN. in *Checking*
	ACKN. spoken by game responder	REPLY-Y	ACKN. spoken by game initiator
		REPLY	

ALIGN moves, the queries which assess the listener's state of understanding and attention, rise in pitch.

It is possible that the high degree of tune variation in ACKNOWLEDGE moves and perhaps in REPLY-Y moves is due to the speaker misinterpreting the written transcript. Recall that reading the transcripts was a swift process, and neither participant had opportunity to rehearse. However, a more likely explanation is similar to that expounded in the discussion of spontaneous results (beginning on page 128). The ACKNOWLEDGE moves might comprise a category in which variation is allowed. Considering this, the whole of the results for read-aloud dialogue reflect favourably upon the hypothesis that discourse function correlates significantly with categories of intonation contour. Some categories of intonation are categories of exclusion.

7.6 Discussion Comparing Spontaneous and Read Results

As hypothesised, utterances in read-aloud dialogue show slightly clearer patterns of correlation between discourse function and intonation contour than they do in spontaneous dialogue. In read dialogue, ALIGN moves have a greater percentage of rising tunes (92%, up from 79%). READY moves consist *only* of low level and falling tunes. ACKNOWLEDGE moves in *Checking*, *Querying-YN*, and *Querying-W* show patterns more skewed toward falling and low level tunes.

A few differences appear between the summary patterns in the read results (Table 7.5) and the spontaneous results (Table 6.11). The read results notably lack a "non-falling" category for tunes. More falling and low level tunes occur in read dialogue. READY moves change from non-rising in spontaneous dialogue to a more uniform falling pattern in read dialogue. ACKNOWLEDGE moves exhibit a greater difference between spontaneous non-falling tunes and read tunes which end low.

It is important to realise that the task of the speaker differs in the spontaneous and read dialogues. In the spontaneous version the task involves helping one participant to draw a path on a map. In the read version, the task involves reading a transcript and successfully re-enacting a dialogue.

One might explain a shift toward falling tunes as a "declarative" strategy in the read-aloud dialogues. The participants are reading a document, the transcript, and may speak the dialogue as they would read a book aloud. There is at least one problem with this, however. The default (or unmarked) Glaswegian intonation contour ends with a nuclear rise. Although the rise may trail off slightly and lower in pitch it remains nonetheless a rise. There is clearly not an increase in rising tunes in the read-aloud dialogues. Also, the dialogues proceed with a remarkable spontaneous quality about them (excerpts from both versions fooled various seminar audiences). At times the speakers may sound slightly bored with the task, but they rarely sound as if they are reading aloud a transcript. Recall that the most "spontaneous sounding" dialogues were chosen for this study.

Some of the tunes differ because utterances were interpreted differently. For example, 16A085, "Yeah", is a REPLY-Y spoken as a question, with a rising tune in the read version. Another example is 35A045, "Mm", a QUERY-W which was interpreted more as an acknowledgement in the read version. It is impossible to read the minds of the speakers to determine what interpretation they assigned to the various conversational moves. One can merely say that overall, the dialogue sounded convincing as a replica of a spontaneous dialogue, and that therefore most of the speakers' interpretations would be similar to those in their original spontaneous dialogues. Clear examples such as the two above are rare.

A shift in intonational tune accompanying a shift in task is not surprising. What is clear from these results is that in both spontaneous and read-aloud dialogue discourse function correlates to intonational tune, and in read dialogue the correlation improves.

Before completing the discussions, it is necessary to examine the data once again, with respect to how individual data points fare across the dialogue conditions.

7.7 Comparison Study: Introduction

In order to more thoroughly understand the spontaneous and read results, we have to assess the shifting of the tune value of individual data points across dialogue conditions. This will help determine whether subjects are implicitly agreeing with the functional distinctions present in the Conversational Games Analysis or whether they are using some other method.

In this study the two sets of data (spontaneous and read) show different tune patterns within each functional category. One might consider the pattern as a histogram showing the frequency of each type of intonation contour for one particular discourse category. Spontaneous results show one histogram for each category, and read results show another. For example, ALIGN moves show a pattern of 19 LH, 3 H, 2 L, 0 HL (and 0 LHL) in the spontaneous condition and 22 LH, 0 H, 1 L, 1 HL, 0 LHL in the read condition. The individual data points must differ in tune value between conditions to create the different histograms (if the histograms do indeed differ). The extent to which the tunes of individual data points differ could be large or small. It is possible that most or all of the data points will change value and still exhibit their respective histograms. It is also possible that very few data points will differ—only enough to achieve the change between histograms. By chance we would expect the individual data points to exhibit a certain level of agreement to achieve the two histograms.

If data points are changing at a "chance" level to create the patterns in the two sets of results, or changing more than one would expect by chance, then we can say that the patterns discovered are valid and the discourse categories we have chosen are appropriate. For example, when a discourse category has non-rising tunes, the tune can freely be any non-rising tune. It is the general pattern of tunes (whether a category of exclusion or inclusion) which holds for a given discourse function, and there is no further specification of discourse function necessary.

If the actual agreement exceeds the level of chance, then subjects are being more consistent than we expect. In other words, they are making better predictions of intonational tune than the Conversational Games Analysis can. They are being more accurate at deciding the tune. If this is the case then it is possible that some other discourse phenomenon is correlating with intonation contour. The speakers would have a definite choice of tune in mind (e.g. high level) but their criteria for choice of tune may be slightly different from the discourse contexts specified in this study. These speakers would be adhering to some as yet unspecified method for determining tune. For example, perhaps their notion of discourse function involves finer distinctions such as separating responses to indirect requests from responses to direct requests.

The hypothesis (as mentioned in the Introduction to the study above, Section 7.1) is that the Conversational Games Analysis adequately represents functional categories and that subjects' behaviour will agree with these categories.

Therefore actual agreement should not exceed the predicted level with any level of significance.

7.7.1 Materials

Materials for the comparison of individual moves across conditions include all of the data points from spontaneous and read dialogues used in the intonation studies above (Chapters 6 and 7).

7.7.2 Method

The statistics described below ascertain the level of chance for individual moves to agree in tune value across dialogue conditions given the different histograms for tunes in the two conditions. As an example, if one discourse context has 4 falling and 4 low level tunes in both conditions, one would expect that some of the moves have held their tune value across conditions. If the moves representing the 4 falling tunes in the spontaneous condition were to shift to the 4 low level tunes in the read condition (and vice-versa), then it would be important to realise that all of the moves have changed their tunes. Similarly, if none of the tunes shift, then it is important to know, as the discourse category may not be narrow enough to map to specific tune categories.

The statistic applied to the two sets of results addresses one discourse context (or function) at a time. The probable (or chance) percentage of agreement for one discourse category is computed as follows (Isard, p.c.):

$$100\times\overset{\textit{Rising}}{\begin{pmatrix} spontRF \\ \times \\ readRF \end{pmatrix}}+\overset{\textit{High Level}}{\begin{pmatrix} spontRF \\ \times \\ readRF \end{pmatrix}}+\overset{\textit{Low Level}}{\begin{pmatrix} spontRF \\ \times \\ readRF \end{pmatrix}}+\overset{\textit{Falling}}{\begin{pmatrix} spontRF \\ \times \\ readRF \end{pmatrix}}+\overset{\textit{Rising-Falling}}{\begin{pmatrix} spontRF \\ \times \\ readRF \end{pmatrix}}$$

= probable percentage of agreement between individual data points across conditions (where spontRF (or readRF) is the relative frequency of spontaneous (or read) data points in the discourse category with a particular tune)

Taking an example from the results, the chance level of agreement for individual data points classed as ALIGN moves would be

$$100 \times \left(\left(\frac{19}{24} \times \frac{22}{24}\right) + \left(\frac{3}{24} \times \frac{0}{24}\right) + \left(\frac{2}{24} \times \frac{1}{24}\right) + \left(\frac{0}{24} \times \frac{1}{24}\right) + \left(\frac{0}{24} \times \frac{0}{24}\right) \right)$$

$$= 72.9\%$$

Therefore it is probable that 72.9% of the ALIGN moves will carry the same tune across speech conditions.

Actual agreement is calculated by simply looking at the results of the two studies, counting the number of data points which have the same tune value across conditions, and finding the percentage of them out of the total number of data points in one condition.

If the percentage of actual agreement for a discourse category is significantly higher than the predicted amount, then speakers are being more consistent than we would expect and are probably using some other notion of discourse function. If the actual agreement falls at or below the predicted amount, then we know that speakers are not differentiating between tune types in that discourse category, and thus, that our discourse categories are acceptable.

Two other values are calculated to put the comparison into perspective: best and worst possible agreement. Worst possible agreement is not necessarily zero. It is calculated for each move by the following method:

> Pick the largest single tune category in either (spontaneous or read) condition. Add up the data points in both conditions which are *not* in that category. Subtract that number from the total number of moves (in one condition, shown in Table 7.6). If the difference is negative, make it zero. Divide the difference by the total number of moves (in one condition) and multiply by 100 to obtain the worst possible agreement.

For ALIGN moves the worst possible agreement is

$$100 \times (24 - (3 + 2 + 1 + 1))\,/24 = 70.8\%$$

Best possible agreement refers to the best level of agreement subjects could possibly reach, given the change in the tune patterns for that functional category. It is calculated by adding together the smaller of the two numbers (from the two conditions) in each tune category, dividing the sum by the total number of data points in one condition, and multiplying by 100. For ALIGN moves, the best percentage of agreement is

$$100 \times (19 + 0 + 1 + 0 + 0) / 24 = 83.3\%$$

Worst and best are presented in the table to provide a scale for the results in each category.

Significance for the agreement levels is calculated with the Kappa statistic (Siegel and Castellan, 1988). The statistic applies to raw data points and provides a slightly different means of computing agreement, similarly adjusting for expected agreement. Instead of testing agreement between subjects, as is the usual application of Kappa, here it tests agreement between individual data points in different dialogue modes.

Kappa values are not presented below (in Table 7.6) because they hide some information displayed in the table and duplicate the rest. For example, Kappa hides the worst and best agreement values inside its formulas.

7.7.3 Results and Discussion

Table 7.6 shows the agreement between individual data points. It includes the expected agreement (or chance agreement), actual agreement, the difference between expected and actual agreement, worst possible agreement, and best possible agreement.

Two of the categories show agreement significantly above chance. One of them we can discount. The higher level of actual agreement for ACKNOWLEDGE moves suggest that speakers might be using some strategy slightly different from function defined as a single move. By including the identity of the game in which the move occurs, discourse function is further specified. The table shows that speakers agree with move x in game y categories except in the case of ACKNOWLEDGE in *Instructing* games. It is possible that in this case speakers have a notion of discourse function that has not been identified to a high enough level of specification. Note however, that even here, the level of actual agreement between data points is only 44% out of a best possible 78%. Speakers are only moderately more accurate at predicting intonational tune than the framework of Conversational Games.

If Actual agreement hovered somewhere near the “best” levels, then there might be some cause for concern that the Conversational Games Analysis is not quite finding the discourse distinctions necessary for a study of intonational behaviour. The fact that they instead hover close to the “expected” levels indicates that subjects are behaving more or less as expected. The Conversational Games Analysis is proven to be a reasonable framework within which to study the behaviour of intonation.

Table 7.6: Agreement Between Tunes Associated with Individual Data Points Across Dialogue Conditions: Expected (Probable) and Actual Agreement, their Difference, Worst and Best Possible Agreement, and the Number of Data Points in One Condition (‡ indicates Kappa value significance in the difference at $p < .01$)

		Percentage Agreement					
Move	*Game*	*Expected*	*Actual*	*Diff.*	*Worst*	*Best*	*Number*
ALIGN		72.9	75.0	2.1	70.8	83.3	24
READY		35.7	28.6	-7.1	21.4	71.4	14
ACKNOWL.‡		26.2	43.8	17.6	0	75.6	160
REPLY-Y		36.0	47.5	11.5	3.3	85.2	61
REPLY-N		41.0	60.0	19.0	20.0	80.0	10
ACKNOWL.	Instructing‡	26.8	44.1	17.3	0	78.4	111
	Explaining	8.3	16.7	8.4	0	16.7	6
	Aligning	100.0	100.0	0	100.0	100.0	2
	Checking	58.3	58.3	0	58.3	58.3	12
	Querying-YN	31.4	36.4	5.0	9.1	59.1	22
	Querying-W	26.5	42.9	16.4	0	57.1	7
REPLY-Y	Querying-YN	37.6	51.7	14.1	6.9	79.3	29
	Checking	36.7	53.8	17.1	7.7	92.3	13
	Aligning	33.2	36.8	3.6	0	73.4	19

Chapter 8

Conclusion

8.1 How the Hypotheses Fared

Two hypotheses were tested in the studies of spontaneous and read dialogue (Chapters 6 and 7; page 100 and page 130):

1. Discourse function will correlate significantly with category of intonation contour.

2. Read speech will show slightly better patterns of correlation than will spontaneous speech.

The hypotheses rely crucially upon the Conversational Games Analysis to adequately represent the function of an utterance in dialogue. Considering that the analysis was developed with various uses in mind, that other projects have used it successfully (e.g. Miller and Weinert, 1995; Newlands *et al.*, 1996; Doherty-Sneddon *et al.*, forthcoming; Anderson *et al.*, forthcoming; Garner *et al.*, 1996; and Taylor *et al.*, 1996), and that coders have high levels of agreement with each other (see experiments in Sections 4.4 and 4.5), it was expected to provide a reliable basis for this study. The comparison of individual data points (Section 7.7) supports the conclusion that the Games Analysis relates closely to categories which speakers implicitly use for their intonation strategies.

8.1.1 Intonation Correlates with Function.

The whole of the results for spontaneous and read-aloud dialogue reflect favourably upon the hypothesis that discourse function correlates significantly with categories of intonation contour. The literature suggests that single intonation categories should map to single discourse categories. However, where one expects only categories of inclusion, the case is often one of exclusion. ALIGN moves display an inclusive category. They rise in tune in both speech modes. Other moves associate with exclusive categories. READY and ACKNOWLEDGE moves spoken before the same person continues talking have non-rising tunes in spontaneous speech and end low in read speech. REPLY moves in both dialogue modes are non-rising (or end low). ACKNOWLEDGE moves in *Instructing* and *Explaining* games (these are primarily Response moves in an Initiation-Response-Feedback structure which do not always control the 'floor') have non-falling tunes in spontaneous speech and non-rising tunes in read speech. ACKNOWLEDGE moves in information-seeking games (usually Feedback moves which control the 'floor') are non-rising in spontaneous speech and end low in read speech.

Individual speaker strategies do not appear to play a significant role overall. Some effects show with ACKNOWLEDGE moves in certain contexts. In *Instructing* games from spontaneous dialogue, two speakers contribute almost equal numbers of rising and falling tunes to a general pattern of non-falling tunes. The other speakers contribute tunes which end low. Speaker strategy appears to account for the random distribution of tunes in the embedded *Querying* games. Three of the speakers contribute non-falling tunes and three contribute non-rising tunes.

Lexical effects do not play a significant role overall, either. In the read dialogue, there may be some effect present in the results of the ACKNOWLEDE moves in *Instructing* games (the only category with a distribution not significantly different from random). The lexical item "mmhmm" correlates with non-falling tunes and contributes 12 of the 34 rises. "Right" is mostly non-rising. These are the only two lexical items which exhibit non-random patterns.

Overall, discourse function correlates with tune categories.

8.1.2 Correlations in Read Dialogue are Better.

As hypothesised (in hypothesis # 2), utterances in read-aloud dialogue show slightly clearer patterns of correlation between most categories of discourse function and intonation contour than utterances in spontaneous dialogue. In read dialogue, ALIGN and READY moves have greater percentages of the most common tunes for those moves, rising and falling, respectively. ACKNOWLEDGE moves in *Checking*, *Querying-YN*, and *Querying-W* games (which are non-rising) show patterns more skewed toward falling tunes in read speech (falling and ending low).

The only pattern which became less clear in the read condition is that for ACKNOWLEDGE moves in *Instructing* games which follow instructions. The pattern was not significantly different from random. This will be considered shortly.

Although patterns in read dialogue are slightly clearer, they still largely pinpoint categories of exclusion rather than categories of inclusion for intonation. That is, one can say that an utterance with a particular function is likely not to have a particular intonational contour. While the tradition in the literature generally is to seek correlation between narrow intonation categories and specific functions (See Section 3.4), in fact it might be a better research strategy to see which intonation categories *never* associate with specific functions. Lisker (p.c.) found categories of exclusion in speech arising from a reading task. His study, in which native speakers read aloud non-English texts several times, found that pitch accent types for particular sentences were generally in free variation, excepting one category. Speakers reading aloud text appear to have rules for intonation assignment which operate on an exclusive-category rather than a single-category basis. It is possible that the published literature is lacking such results because authors consider a failure to find a correlation with a single category to be a failure altogether. While it is obvious that there cannot be one-to-one mappings of intonation contour to meaning (our shades of meaning would be limited to the number of different contours), it is not so obvious that a *lack* of correlation between a functional category and a particular intonational category might be an appropriate way to orient one's view. That is, it might be more appropriate to identify intonation tune x that never is used in a particular context. If the read speech results are pinpointing clearer correlations, then it is possible that something like "ends low" is a more real category than low level (L) or fall (HL).

If one works from the premise that categories of exclusion are as valid as categories of inclusion, then the problem of understanding the "messy" categories from the data may become a less daunting task.

Consider the "messiest" category from the results, that of ACKNOWLEDGE moves. In some discourse contexts, this type of move appears to associate with all intonation contours equally. There are a number of possible explanations, including the following three. Firstly, the ACKNOWLEDGE moves might comprise a category in which variation is allowed. There might be no intonational difference in meaning, and the speaker might change contour type to avoid being a boring speaker. Secondly, there could be some difference in function which is important but we haven't identified. Thirdly, it is possible that the speaker changes tune to indicate some meaning to the hearer which is superfluous to the functional interpretation of the utterance and the shared linguistic plan for the conversation. The third is the preferred explanation. Results from the Kappa test (Section 7.7) indicate that some factor may be involved in choosing intonation contours for ACKNOWLEDGE moves after INSTRUCT moves in *Instructing* games that we have not identified.[1]

Based upon examination of the dialogues (and videos) from the Map Task Corpus, especially the ones included in the present study, it appears as though some of the acknowledgements spoken indicate a speaker's attitude toward the

[1] Speakers in this context are moderately more accurate at predicting intonational tune than the framework of Conversational Games (44% consistent out of a best possible 78%).

task, or level of confidence, rather than any information relevant to the procedure of the conversation. When a speaker is momentarily bored, lost with regard to the route, or so deeply engaged in activity related to the task that their attention shifts away from the other person, their pitch range appears to reduce dramatically and level tunes emerge in the brief utterances. A rise in tune may indicate a greater level of enthusiasm and energy on the part of the speaker. Alternatively, it is possible that the rise versus fall distinction also indicates a speaker's internal (mental) state with respect to underlying goals which conflict with the goals of the other person. For instance, the speaker may have a very short-term goal such as drawing one small segment of a larger line (which might be more relevant to the information giver's goal) or waiting to hear part of an answer to a question. It is possible that if the short-term goal is completed, the speaker would answer with one tune and if the goal is on-going, the speaker would answer with the other tune. Unfortunately, it is difficult for the analyst to delve that deeply into the mind of the speaker.

8.2 Differences Between Spontaneous and Read Speech

A few differences appear between the summary patterns in the spontaneous and read results (Tables 6.11 and 7.5). The read results notably lack a "non-falling" category for tunes. More falling and low level tunes occur within most discourse categories in read dialogue. Ready moves change from non-rising in spontaneous dialogue to a falling pattern in read dialogue. Acknowledge moves overall exhibit a preference for non-falling tunes in spontaneous dialogue and tunes which end low in read dialogue.

Although some overall differences in tune strategy appear clearly, what is not clear from the results is what that difference represents. The speakers in read dialogue may aim for something similar to speakers in a spontaneous dialogue situation as their strategy. In other words, the change may reflect a better surfacing of underlying intonation categories ("better representation" strategy). This was the assumed strategy when the hypothesis was presented (Section 7.1). On the other hand, the task of the speaker actually differs in the spontaneous and read dialogues. In the spontaneous version the task involves helping one participant to draw a path on a map. In the read version, the task involves reading a transcript and successfully re-enacting a dialogue. The strategy of the speakers in read dialogue may reflect the difference in task ("task-based" strategy). One notable difference bewteen the dialogue modes relates to turn-taking. In read dialogue turn-taking cues are not required. Readers try to re-enact the original situation but may miss this aspect of spontaneous conversation. They might not necessarily imitate turn-taking interactive style. Identifying the source of strategy is made more difficult because the categories are ones of exclusion, and the changes across speech modes are small.

In fact, it appears as though both strategies are in action. The results show that in the Glasgow accent, non-rising tunes and tunes ending low are continuation markers in spontaneous and read speech, respectively (it is in READY moves and ACKNOWLEDGE moves which precede the same speaker's continuation of turn). In ACKNOWLEDGE moves which control the 'floor' in information-seeking games, tunes are generally non-rising in spontaneous speech and end low in read speech. This indicates that speakers may aim for an 'ends low' continuation marker to keep the 'floor'. In ACKNOWLEDGE moves in information-giving games, the speaker relinquishes the 'floor' much of the time. Here, the spontaneous tune is non-falling, indicating that it is *not* a continuation marker. However, the read speech has a non-rising strategy. The speakers in the reading task may treat all utterances like floor-holding situations because in a reading task the speaker has his or her own (floor-holding) turn as specified in the transcript, even in contexts which correspond to non-floor-holding turns in spontaneous dialogue. This shows that turn-taking is being handled differently (supporting the "task-based" strategy). Speakers appear to be producing a "better representation" strategy as well, as one can see by the 'ends low' continuation strategy in read speech compared to the 'non-rising' continuation strategy in spontaneous speech.

Other studies (e.g. Blaauw, 1995) have found a marked difference between intonation in spontaneous and read materials (monologue), noting that more falling boundary tones are found in read speech. This suggests some sort of "declarative" strategy in the read-aloud dialogues if the same principle is working. However, the speakers in the read dialogues perform remarkably well at producing a spontaneous quality. (Excerpts from both versions fooled various seminar audiences.) The speakers occasionally sound slightly bored with the task, but they rarely sound as if they are reading aloud a transcript.

The problem of fleshing out finer details of the source of speaker strategy across dialogue modes is left for future research (see below).

8.3 Implications for the Use of Read Speech

One may ask if read dialogue is perhaps adequate for training intonation models in speech recognisers. The results of the intonation studies suggest that it is probably not. Read dialogue does not offer the same variety of intonation that spontaneous speech does, and strategies appear to differ slightly. If one is willing to accept a certain level of 'error' at the outset, then read-aloud dialogue may prove an acceptable substitute for eliciting the same intonation strategies that are found in spontaneous dialogue.

8.4 Future Directions

In order to fully understand the details of the intonation strategies in spontaneous and read dialogue, further work should be undertaken. The balance between a "better representation" strategy and a "task-based" strategy in read speech relates to the intent of the speaker – not always an easy problem to solve. One could approach the problem by recording a new set of dialogues, or excerpts of dialogues, and interviewing the speakers immediately afterwards. The speakers can, at least, indicate whether they thought they were interpreting phrases correctly or rate themselves on a scale – something that would support the "better representation" interpretation. Small studies could be organised around getting speakers to re-enact turn-taking situations better. Rehearsals might help speakers to imitate these excerpts.

Perceptive studies could lend to our understanding of how categories of exclusion operate. We could test the "natural" quality of various tunes in utterances by devising experiments which involve listener judgements.

Finally, it is worth examining the intonation function of longer utterances in dialogue by extending the studies in Chapters 6 and 7. Grice *et al.* (1995) have done something like this for CHECK and QUERY moves in three European languages.

Thorough empirical studies in these directions would add to our understanding of how intonation functions in dialogue and thus help us identify the nature of the way in which people carry across meaning in conversation.

Appendix A

Example of a Coded Dialogue

The following dialogue was used in the intonation studies of Chapters 6 and 7. It comes from the dialogues in the Map Task Corpus (introduced in Section 4.2) where there was no eye contact. The two participants in this dialogue are unfamiliar with each other. Motivation for the choice of dialogue can be found in Section 6.2.

Two types of notation are present in the dialogue below, transcription notation and game coding.

Transcribers added several symbols to the basic orthographic transcription. Information concerning the whole dialogue is noted with a starred D (∗D). Turns begin with a starred T immediately followed by a letter indicating instruction giver (∗TA) or follower (∗TB). The turn symbols are followed by the sequential number of turn in the dialogue. Angled brackets (⟨ ⟩) mark the region in which one speaker overlaps the other. The point at which the interruption occurs is marked with an oblique line (/). Curly brackets ({ }) indicate a word or fragment outside a typical lexicon. The left curly bracket is immediately followed by a single letter code which indicates what type of non-word follows. There are six codes used in curly brackets:

g	'grunt' (e.g. in turn *TB6)
a	abandoned word (e.g. in turn *TB12)
m	filled pause (e.g. in turn *TB16)
x	cross between filled pause and 'grunt' (e.g. in turn *TB32)
i	initial partial word; intentional miss of first syllable (e.g. in turn *TA37)
c	cited word; for feature names printed on maps (e.g. in turn *TA59)

The orthography of the initial partial word is immediately followed by an equal sign (=) and a guess of the full word, e.g. {i cause=because}.

Game coding has two components, moves and games. Moves are noted with a starred M (*M) followed by the name of the move or moves and any features (detailed in Section 4.3.4, beginning page 55) immediately beneath the orthography representing the utterance which they code. Square brackets ([ ]) enclose comments about moves. Beginnings of games are noted with a starred E followed by the sequential number of the game, a code for information giver (IG) or follower (IF) who initiates the game, the name of the move initiating the game, and optionally, a code indicating that the game is embedded. So, for example, the line "*E 3 IF query-yn em" shows that the third game of the dialogue, an embedded *Querying-YN* game initiated by the information follower, is about to begin. Game endings are simply noted with starred end symbol (*End) followed by the number of the game which has just concluded.

*D Eye-Contact: NO

*D Quad: 1

*D Number: 6

*D Identifier: naq1c6–4du

*D Instruction-Follower: Eileen

*D Instruction-Giver: Mark

*E 1 IG explain

*TA 1

Okay,

*M ready

the start's at the top left.

*M explain

*TB 2

Right, aye, I've got the start marked down.

*M acknowledge

*End 1

*E 2 IG query-yn

*TA 3

You have cliffs there?

*M query-yn

*E 3 IF query-yn em

*TB 4

Sandstone cliffs?

*M query-yn

*TA 5

Yeah.

*M reply-y

*End 3

*TB 6

{g Mmhmm}.

*M reply-y

*End 2

*E 4 IG query-yn

*TA 7

You don't have a forge, do you?

*M query-yn

*TB 8

No.

*M reply-n [agrees]

*End 4

*E 5 IG explain

*TA 9

Right,

*M ready

there's a forge about two inches beneath the cliffs.

*M explain

*E 6 IG align em

Okay?

*M align

*End 6

*TB 10

Right,

*M acknowledge

*End 5

*E 7 IF query-yn

directly down?

*M query-yn

*TA 11

Yeah.

*M reply-y

∗End 7

∗E 8 IF query-yn

∗TB 12

{a Besi} beside an old pine?

∗M query-yn

∗TA 13

Yeah,

∗M reply-y

it's about, you know, an inch or so, two inches from ... to the left of the old pine.

∗M clarify

∗End 8

∗E 9 IG instruct

So if you just take a line straight down from the start.

∗M instruct

∗TB 14

⟨ {g Uh-huh}.

∗M acknowledge

∗E 10 IG align em

∗TA 15

Okay?

∗M align

∗End 10

Four inches down /

∗M instruct cont

∗E 11 IF align em

∗TB 16

So ... hang on ... You see this {m ehm},

∗M align aban

∗End 11

∗TA 17

four inches down, just take a straight line. ⟩

∗M instruct cont

∗E 12 IF query-yn em

∗TB 18

Well, is this is this next to the ... is it beside the old pine? Sort of or is it on a sort of

∗M query-yn

∗TA 19

⟨ No,

∗M reply-n

just draw a straight line down from the start, four inches or so, first /

∗M clarify

∗TB 20

Right.

∗M acknowledge

∗TA 21

of all,

∗E 13 IG align em

okay.

∗M align

*End 13

*End 12

*E 14 IG explain em

You don't have a forge there, but you know that's that's beneath /

*M explain

*TB 22

Where it is,

*M fill

*TA 23

the forge.

*M explain cont

*E 15 IG align em

Okay. ⟩

*M align

*End 15

*TB 24

Right.

*M acknowledge

*End 14

*End 9

*E 16 IG instruct

*TA 25

Then go right, to the bottom of the old pine.

*M instruct

*TB 26

{g Mmhmm}.

∗M acknowledge

∗End 16

∗E 17 IG instruct

∗TA 27

Go up, and round the old pine.

∗M instruct

∗TB 28

{g Mmhmm}.

∗M acknowledge

∗End 17

∗E 18 IG check

∗TA 29

You've got a bay at the top, haven't you?

∗M check

∗E 19 IF check em

∗TB 30

A bay? A green bay,

∗M check

∗End 19

{g uh-huh}.

∗M reply-y

∗End 18

∗E 20 IG explain

∗TA 31

I don't have a bay there, but

∗M explain

∗TB 32

{x Oh}, right.

∗M acknowledge

∗End 20

∗E 21 IG instruct

∗TA 33

If you just get to the ... so, you know, if you just go round the old pine,

∗M instruct

∗E 22 IG align em

okay?

∗M align

∗End 22

∗TB 34

{g Uh-huh}.

∗M acknowledge

∗End 21

∗E 23 IG instruct

∗TA 35

Once you get to the top,

∗M instruct

∗TB 36

{g Mmhmm}.

∗M acknowledge

∗TA 37

go two or three inches to the right {i til=until} you get to the pine forest,

*M instruct cont

*E 24 IG align em

okay?

*M align

*End 24

*TB 38

{g Mmhmm}.

*M acknowledge

*End 23

*E 25 IF explain

Round to green bay.

*M explain mumbl

*End 25

*E 26 IG instruct

*TA 39

Go down on the left-hand side of the pine forest,

*M instruct

*TB 40

The left-hand side, right.

*M acknowledge repo

*TA 41

{g Mmhmm},

*M acknowledge

*TB 42

{g Mmhmm}.

∗M acknowledge

∗End 26

∗E 27 IG instruct

∗TA 43

draw a line straight down,

∗M instruct

∗TB 44

{g Mmhmm}.

∗M acknowledge

∗TA 45

until you get to just about a centimetre or or two northwest of the bakery,

∗M instruct cont

∗E 28 IG align em

okay?

∗M align

∗End 28

In other words, more or less the top left-hand side of the bakery, but a an inch or so above,

∗M clarify

∗TB 46

Right. {g Mmhmm}.

∗M acknowledge

∗End 27

∗E 29 IG instruct

∗TA 47

So {a th} ... Then you go just go round the bakery, in an oval shape,

*M instruct

*E 30 IG align em

okay,

*M align

*End 30

*TB 48

Right.

*M acknowledge

*TA 49

right round it on the right-hand side.

*M instruct cont

*TB 50

Right.

*M acknowledge

*End 29

*E 31 IG query-yn

*TA 51

And you don't have a canal, do you?

*M query-yn

*TB 52

No.

*M reply-n [agrees]

*End 31

*E 32 IG explain

∗TA 53

The canal's about two inches to the left of the bakery,

∗M explain

∗End 32

∗E 33 IG instruct

{a bu} ... so, go round the bakery,

∗M instruct

∗TB 54

{g Mmhmm}.

∗M acknowledge

∗TA 55

and stop about an inch to the left of the bakery.

∗M instruct cont

∗TB 56

{g Mmhmm}.

∗M acknowledge

∗End 33

∗E 34 IG instruct

∗TA 57

Then draw a line straight down

∗M instruct

∗TB 58

{g Mmhmm}.

∗M acknowledge

∗TA 59

To where crane bay starts to curve above {a abo} above the gap, but it's between {c crane} and the word {c bay}.

*M instruct cont

*TB 60

⟨ Right.

*M acknowledge

*E 35 align em

*TA 61

Okay? ⟩

*M align

*TB 62

{g Mmhmm}.

*M reply-y

*End 35

*End 34

*E 36 IG instruct

*TA 63

So draw a line straight down to there, follow that curve to the left,

*M instruct

*TB 64

{g Mmhmm}.

*M acknowledge

*TA 65

⟨ until crane bay curves quite steeply away,

*M instruct cont

*TB 66

Down. ⟩

*M fill

{g Mmhmm}.

*M acknowledge

*TA 67

But don't go don't go round there,

*M instruct cont

*E 37 IG align em

okay.

*M align

*End 37

*TB 68

Right, {g mmhmm}.

*M acknowledge

*End 36

*E 38 IG query-yn

*TA 69

So then, do you have wheatfields down there, next to the bay?

*M query-yn

*TB 70

No,

*M reply-n

I've got them up beside the old pine.

*M clarify

∗TA 71

Yeah,

∗M acknowledge

but you don't have to worry about those, right now.

∗M clarify

∗TB 72

⟨ No. Don't have them /

∗M acknowledge reps

∗TA 73

Okay then.

∗M acknowledge

∗End 38

∗E 39 IG explain

I've got wheatfields in line /

∗M explain aban

∗End 39

∗E 40 IF query-yn

∗TB 74

Have you got a rocket warehouse? ⟩

∗M query-yn

∗TA 75

⟨ Yeah,

∗M reply-y

∗End 40

∗E 41 IG explain

I've got wheatfields in line with the word {c bay}, about, /

∗M explain

∗TB 76

{g Mmhmm}.

∗M acknowledge

∗TA 77

say, an inch to the left-hand side of the bay. ⟩

∗M explain cont

∗TB 78

{g Mmhmm}.

∗M acknowledge

∗End 41

∗E 42 IG query-yn

∗TA 79

So, if ... {a wh} where are you now? The top left-hand side of crane bay?

∗M query-yn

∗TB 80

{g Uh-huh}.

∗M reply-y

∗End 42

∗E 43 IG instruct

∗TA 81

Just draw a line two inches down. Or ... in fact just draw a line down until an inch or so above the rocket warehouse.

∗M instruct

*TB 82

Right.

*M acknowledge

*E 44 IG align em

*TA 83

Okay?

*M align

*End 44

*End 43

*E 45 IG instruct

⟨ Go round the rocket warehouse in the same way as you went round the bakery, but {a o} {a o} ... to the left-hand side obviously /

*M instruct

*E 46 IF check em

*TB 84

To the left-hand side?

*M check

Right, {g uh-huh}.

*M acknowledge

*E 47 IG query-yn em

*TA 85

You're towards the outside of the page?

*M query-yn

*End 47

Yeah. ⟩

*M reply-y

*End 46

*End 45

*E 48 IG instruct

And, so if you've gone round it ... {a don} don't go beneath it though,

*M instruct

*TB 86

Right.

*M acknowledge

*E 49 IG align em

*TA 87

Okay?

*M align

*TB 88

{g Uh-huh}.

*M reply-y

*End 49

*End 48

*E 50 IG instruct

*TA 89

⟨ Once you get to the bottom of it, on the left-hand side of the rocket warehouse,

*M instruct

*E 51 IG align em

okay /

*M align

*End 51

*TB 90

{g Uh-huh}, {g uh-huh}. ⟩

*M acknowledge

*TA 91

⟨ {x Eh}, if you just draw a line more or less straight down to the left-hand side of the lighthouse because that's where the cross is for the finish. You've got a /

*M instruct cont

*TB 92

What {a a}

*E 52 IG query-yn em

*TA 93

lighthouse, haven't you?

*M query-yn

*End 52

*E 53 IF query-w em

*TB 94

What about the {a ca} there's a cave. ⟩

*M query-w

*E 54 IG query-w em

*TA 95

Where's the cave?

*M query-w

*E 55 IF align em

*TB 96

Right,

∗M ready

see the rocket warehouse and the old lighthouse?

∗M align

∗TA 97

Yeah.

∗M reply-y

∗End 55

∗TB 98

⟨ Right between that to the left, it's sort of like in /

∗M reply-w

∗TA 99

Okay, right,

∗M acknowledge

∗TB 100

a triangle. ⟩

∗TA 101

Okay.

∗M acknowledge

∗End 54

∗TB 102

Do you have to go round the cave or?

∗M query-w cont

∗TA 103

I don't ... I don't have a cave.

*M explain
*End 53
*End 50
*E 56 IG instruct
{a T} take the line from the rocket warehouse,
*M instruct
*E 57 IG align em
right,
*M align
*End 57

*TB 104
{g Uh-huh}.
*M acknowledge

*TA 105
don't go straight down, take it right to the left-hand side of the
page.
*M instruct cont

*TB 106
To the left-hand side? Right,
*M acknowledge repo
*E 58 IF query-yn em
so it will be round the cave?
*M query-yn

*TA 107
Almost right next to it.
*M reply-w

*TB 108

Right.

*M acknowledge

*End 58

*End 56

*E 59 IG instruct

*TA 109

⟨ And only curve in where the bay starts starts again then, trying you know,

*M instruct

*E 60 IG align em

right? /

*M align

*End 60

*TB 110

The word {c bay} starts again?

*M acknowledge repo

*E 61 IF query-yn em

Do you not right sort /

*M query-yn aban

*End 61

*E 62 IG align em

*TA 111

See the see the ... see the old lighthouse is? ⟩

*M align

*TB 112

{g Mmhmm}.

*M reply-y

∗TA 113

Okay.

∗M acknowledge

∗End 62

∗E 63 IF query-yn em

∗TB 114

Just just draw a line straight through there?

∗M query-yn

∗TA 115

⟨ Yeah.

∗M reply-y

Draw draw /

∗M clarify

∗E 64 IF query-yn em

∗TB 116

In a sort of curve?

∗M query-yn

∗TA 117

the line ... yeah,

∗M reply-y

draw the line down to where where the land ends,

kind of thing. ⟩

∗M clarify cont

∗TB 118

Right, {g uh-huh}.

∗M acknowledge

∗TA 119

⟨ And then ... so /

∗TB 120

{x Oh}, right.

∗M acknowledge

∗End 64

∗End 63

∗End 59

∗E 65 IG instruct

∗TA 121

then then {a cu} ... you know, draw a line to the right-hand side ...

sorry, to the right. ⟩

∗M instruct

∗TB 122

Right.

∗M acknowledge

∗TA 123

⟨ to where the where the cross is.

∗M instruct cont

∗TB 124

{i Til=Until} ... ⟩

∗M mumbl

∗TA 125

*E 66 IG query-yn em

You have a lighthouse don't you?

*M query-yn

*TB 126

{g Uh-huh}.

*M reply-y

*End 66

*E 67 IG explain em

*TA 127

The cross is just next to the lighthouse.

*M explain

*TB 128

Right,

*M acknowledge

*End 67

*E 68 IF query-w em

which side?

*M query-w

*TA 129

The left-hand side, sorry.

*M reply-w

*TB 130

Right.

*M acknowledge

*End 68

∗E 69 IG align em

∗TA 131

Okay?

∗M align

∗TB 132

{g Mmhmm}.

∗M reply-y

∗End 69

∗End 65

∗E 70 IG explain

∗TA 133

That's it.

∗M explain

∗End 70

Appendix B

Sample Raw Results from the Studies

This section contains two excerpts of the raw results as preserved in the summary data file, showing tunes associated with each of the data points, organised by conversational move. The first excerpt shows data from spontaneous dialogue and the second from read-aloud dialogue. (Note that accent targets and their corresponding F0 points were often taken at more than one point for level portions of tunes.)

Sample of Results from Spontaneous Dialogue

NAQ1C6 SPEAKER A, INFORMATION GIVER, Mark
NAQ1C6 SPEAKER B, INFORMATION FOLLOWER, Eileen

For move coding, UPPERCASE indicates utterances spoken by the other person. Under PREV-MOVE, the combination "*move1) move2*" stands for the previous two moves—move1 which ended another game and move2, the immediately previous move within the current game. A single right parenthesis, ")", means end of game.

Under NEXT-MOVE, an utterance by the other speaker is only mentioned if there is no end of game (")") or utterance by the same speaker in that turn.

```
UTT#     MOVE      GAME      PREV-MOVE NEXT-MOVE-
UTT#     ORTHOGRAPHY
UTT#     H/L       -F0-      [comments]

------------------------------------------------------------
16A009   align     al em     ex        )
16A009   (Right, there's a forge about two inches beneath
16A009   the cliffs.) Okay?
16A009   L         146       [separate phrases]
16A009   H         159
------------------------------------------------------------
16A015   align     al em     AC        ) in cont
16A015   Okay? (Four inches down)
16A015   L         139       [sep phrases; interrupts speaker]
16A015   H         157
------------------------------------------------------------
16A021   align     al em     AC        ) ex
16A021   (of all,) okay. (You don't have a forge there,
16A021   but w you know that's that's beneath)
16A021   L         142       [sep phrases]
16A021   H         153
------------------------------------------------------------
16A023   align     al em     ex cont   )
16A023   (the forge.) Okay.
16A023   L         141       [sep phrases; looks level, sounds rising]
16A023   H         147       [sounds LH because 'forge okay' is LH]
------------------------------------------------------------
16A033   align     al em     in        )
16A033   (If you just get to the ... so, you know, if you
16A033   just go round the old pine,) okay?
16A033   L         143       [sep phrases]
```

```
16A033   H        150
------------------------------------------------------------
16A045   align    al em    in cont    ) cl
16A045   (until you get to just about a centimetre or or
16A045   two northwest of the bakery,) okay? (In other
16A045   words it, more or less the top left-hand side of the
16A045   bakery, but a an inch or so above,)
16A045   L        127      [sep phrases]
16A045   H        137      [might remotely be level]
------------------------------------------------------------
16A047   align    al em    in         )
16A047   (So th... Then you go just go round the bakery,
16A047   in an oval shape,) okay,
16A047   L        136      [separate phrases; maintained H at end]
16A047   H        153
------------------------------------------------------------
16A061   align    al em    AC         R-Y
16A061   Okay?
16A061   L        144      [interrupts speaker]
16A061   H        154      [sounds almost level]
16A061   (M)      145
------------------------------------------------------------
16A067   align    al em    in cont    )
16A067   (But don't go don't go round there,) okay.
16A067   L        151      [separate phrases]
16A067   H        163
16A067   (M)      153
------------------------------------------------------------
16A083   align    al em    AC         ) in
16A083   Okay?
```

```
16A083   (Go round the rocket warehouse in the same
16A083   way as you went round the bakery, but o o ...
16A083   to the left-hand side obviously)
16A083   L          134        [nasalised 'nkay 1st syll might be spurious]
16A083   H          147        [phrase on its own]
————————————————————————————————————————————————————————————————————
16A087   align      al em      AC         R-Y
16A087   Okay?
16A087   L          145
16A087   L          149
————————————————————————————————————————————————————————————————————
16A089   align      al em      in         )
16A089   (Once you get to the bottom of it, on the left-hand
16A089   side of the rocket warehouse,) okay
16A089   H          144        [all almost monotone; hard to detect any change]
16A089   H          148        [sep phrases]
————————————————————————————————————————————————————————————————————
16A103   align      al em      in         )
16A103   (I don't ... I don't have a cave. T take the line
16A103   from the rocket warehouse,) right,
16A103   H          155        [sep phrases; upside down smiley]
————————————————————————————————————————————————————————————————————
16A109   align      al em      in         )
16A109   (And only curve in where the bay starts starts
16A109   again then, trying you know,) right?
16A109   L          146        [sep phrases]
16A109   H          156
16A109   (M)        147
————————————————————————————————————————————————————————————————————
16A131   align      al em      AC         R-Y
```

16A131	Okay?			
16A131	L	142		
16A131	H	172		

31A007	align	al em	in	AC repo
31A007	Okay?			
31A007	L	224		
31A007	L	233		
31A007	H	262 rise		

32A094	align	al em	AC	CH
32A094	Right?			
32A094	L	204		
32A094	H	223 rise		

Sample of Results from Read Dialogue

NAQ1C6 SPEAKER A, INFORMATION GIVER, Mark

NAQ1C6 SPEAKER B, INFORMATION FOLLOWER, Eileen

UTT#	H/L	-F0-	[comments]
r16A005	L	134	
r16A005	L	131	
r16A009	L	143	
r16A009	H	155	

r16A011	L	134	
r16A011	L	129	
r16A015	L	145	[2nd syll rises in itself; SEP phrases]
r16A015	L	161	
r16A015	H	174	
r16A021	L	137	[2nd syll rises in itself; sep phrases]
r16A021	L	164	
r16A021	H	180	
r16A023	L	138	[2nd syll rises in itself; SEP phrases]
r16A023	L	155	
r16A023	H	179	
r16A033	L	135	[sep phrases]
r16A033	L	160	[2nd syll rises in itself]
r16A033	H	177	
r16A041	H	159	[not sure if drop is due to velar influence]
r16A041	H	148	
r16A045	L	145	[separate phrase; LH upstepped]
r16A045	H	161	
r16A047	L	144	[2nd syll rises as well; sep phrases]
r16A047	H	169	
r16A061	L	148	[very light breath; uncertain F0]
r16A061	L	170	[definitely LH]

r16A061	H	179	
r16A067	L	151	[SEP phrases; text differs from spont. but OK]
r16A067	H	166	[LH(M) is really LH]
r16A067	M	157	
r16A083	L	146	[spurious point, whispered]
r16A083	L	160	[definitely LH]
r16A083	H	168	
r16A085	L	138	[sep phrase; like a question]
r16A085	H	165	
r16A087	L	142	[whispered syll]
r16A087	L	170	
r16A087	H	186	
r16A089	L	141	[all almost monotone; hard to detect any change]
r16A089	H	150	[sep phrases]
r16A097	L	132	
r16A103	L	143	[sep phrases]
r16A103	H	157	
r16A109	L	145	[monotone; text differs from spont.]
r16A109	H	156	[upside down smiley]
r16A109	(L)	136	[sep phrases]
r16A115	H	176	[sep phrases]

r16A115	L	144	
r16A117	L	179	[sep phrases]
r16A117	L	171	
r16A131	L	148	
r16A131	H	179	
r16B006	L	180	
r16B006	L	183	[octave error]
r16B008	L	183	[sort of smiley shape]

References

Allen, James F. and Lenhart K. Schubert (1991) "The TRAINS Project," TRAINS Technical Note 91-1, May, University of Rochester, Computer Science Department, New York.

Anderson, Anne H., Miles Bader, Ellen G. Bard, Elizabeth H. Boyle, Gwyneth M. Doherty, Simon C. Garrod, Stephen D. Isard, Jacqueline C. Kowtko, Jan M. McAllister, Jim Miller, Catherine F. Sotillo, Henry S. Thompson, and Regina Weinert (1991) "The HCRC Map Task Corpus," *Language and Speech*, 34(4):351-366.

Anderson, A. H., A. Robertson, K. Kilborn, S. Beeke, and E. Dean (forthcoming) "Dialogue despite difficulties: A study of communication between aphasic and unimpaired speakers," in T. Givon (ed.) *Conversation*, Amsterdam: John Benjamins, pp. 1-41.

Anderson, AH, A Clark, and J. Mullin (1991) "Introducing information in dialogues: How young speakers refer and how young listeners respond," *Journal of Child Language*, **18**: 663-687.

Armstrong, Lilias E. and Ida C. Ward (1931) *A Handbook of English Intonation*, 2nd Edition, Cambridge: Heffer.

Austin, John Langshaw (1962) *How to do things with words*, Oxford University Press.

Ayers, Gayle M. (1994) "Discourse functions of pitch range in spontaneous and read speech," *Ohio State University Working Papers in Linguistics No. 44*, Linguistics Laboratory, OSU, Ohio, pp. 1-49.

Beattie, Geoffrey W., Anne Cutler, and Mark Pearson (1982) "Why is Mrs. Thatcher interrupted so often?" *Nature* Vol. 300, pp. 744-747, 23/30 December.

Beckman, Mary E. (1986) *Stress and Non-Stress Accent*, Dordrecht: Foris.

Beckman and Ayers (1994) "Guidelines for ToBI Labeling," Department of Linguistics, Ohio State University, Ohio.

Beckman and Pierrehumbert (1986) "Intonation structure in Japanese and English," *Phonology Yearbook* **3**: 255-309.

Bernstein, Jared and Gay Baldwin (1985) "Spontaneous vs. Prepared Speech," Paper Presented at the 110th Meeting of the Acoustical Society of America, Nashville, November.

Blaauw, Eleonora (1994) "The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech," *Speech Communication* **14**: 359-375.

Blaauw, Eleonora (1995) *On the Perceptual Classification of Spontaneous and Read Speech*, Ph.D. thesis, Utrecht: OTS (Research Institute for Language and Speech) Dissertation Series.

Bolinger, Dwight L. (1958) "A Theory of Pitch Accent in English," *Word*, **14**: 109-149.

Bolinger, Dwight (1985) *Intonation and its Parts: Melody in Spoken English*, London: Edward Arnold.

Bolinger, Dwight (1989) *Intonation and its Uses: Melody in Grammar and Discourse*, London: Edward Arnold.

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

Brazil, David (1975) *Discourse Intonation*, Birmingham: University of Birmingham.

Brown, Gillian, Anne H. Anderson, Richard Shillcock, and George Yule (1984) *Teaching Talk*, Cambridge University Press.

Brown, Gillian, Karen L. Currie, and Joanne Kenworthy (1980) *Questions of Intonation*, London: Croom Helm.

Bruce, G. (1982) "Textual Aspects of Prosody in Swedish," *Phonetica* **39**: 274-287.

Bruce, Gösta and Paul Touati (1992) "On the Analysis of Prosody in Spontaneous Speech with Exemplification from Swedish and French," *Speech Communication* **11**: 453-458.

Butler, Charles (1634) *The English Grammar*, London. Reprinted as *Charles Butler's English Grammar*, (ed.) A. Eichler, Halle: Niemeyer, 1910.

Carletta, Jean, Amy Isard, Stephen Isard, Jacqueline Kowtko, Gwyneth Doherty-Sneddon, and Anne Anderson (1995) "The Coding of Dialogue Structure in a Corpus," *Proceedings of the Ninth Twente Workshop on Language Technology: Corpus-Based Approaches to Dialogue Modelling*, (eds.) J.A. Andernach, S.P. van de Burgt, and G.F. van der Hoeven, Universiteit Twente, Enschede, Netherlands, June, pp. 25-34.

Carletta, Jean, Amy Isard, Stephen Isard, Jacqueline Kowtko, Alison Newlands, Gwyneth Doherty-Sneddon, and Anne Anderson (forthcoming), "The Reliability of a Dialogue Structure Coding Scheme," *Computational Linguistics Special Issue*.

Clark, Herbert H. and Edward F. Schaefer, (1987) "Collaborating on contributions to conversations," *Language and Cognitive Processes*, **2**: 19-41.

Clark, Herbert H. (1996) *Using Language*, Cambridge University Press.

Collier, R. (1993) "On the communicative functions of prosody: Some experiments," *IPO Annual Progress Report*, Eindhoven, Netherlands, **28**: 67-75.

Connell, Bruce and D. Robert Ladd (1990) "Aspects of Pitch Realisation in Yoruba," *Phonology* **7**: 1-29.

Coulthard, Malcolm (1985) *An Introduction to Discourse Analysis*, London: Longman.

Cruttenden, Alan (1986) *Intonation*, Cambridge University Press.

Cruttenden, Alan (1995) "Rises in English," in Jack Windsor Lewis (ed.) *Studies in General and English Phonetics: Essays in Honour of Professor J.D. O'Connor*, London: Routledge, pp. 155-173.

Crystal, David (1969) *Prosodic systems and intonation in English*, Cambridge University Press.

Cutler, Anne and Mark Pearson (1986) "On the analysis of prosodic turn-taking cues" in C. Johns-Lewis (ed.) *Intonation in Discourse*, London: Croom Helm, pp. 139-155.

Doherty-Sneddon, Gwyneth (personal communication) University of Sterling, UK.

Doherty-Sneddon, G., A. H. Anderson, C. O'Malley, S. Langton, S. Garrod, and V. Bruce (forthcoming) "Face-to-Face and Video-Mediated Communication: A Comparison of Dialogue Structure and Task Performance," *Journal of Experimental Psychology: Applied.*

French, Peter and John Local (1986) "Prosodic Features and the Management of Interruptions," in C. Johns-Lewis (ed.) *Intonation in Discourse*, London:

Croom Helm, pp.157-180.

Garner, P. N., S. R. Browning, R. K. Moore, M. J. Russell (1996) "A theory of word frequencies and its application to dialogue move recognition," International Conference on Spoken Language Processing, Philadelpia, October.

Garrod, Simon and Anthony Anderson (1987) "Saying what you mean in dialogue: A study in conceptual and semantic co-ordination," *Cognition* **27**: 181-218.

Giles, Howard, Nikolas Coupland, and Justine Coupland (1991) "Accommodation theory: Communication, context, and consequence," in Giles, H., J. Coupland, and N. Coupland (eds.) *Contexts of Accommodation: Developments in Applied Sociolinguistics*, Cambridge: Cambridge University Press, Chapter 1, pp.1-68.

Goldman-Eisler, Frieda (1968) *Psycholinguistics: Experiments in spontaneous speech*, London: Academic Press.

Goldman-Eisler, Frieda (1961) "The Distribution of Pause Durations in Speech," *Language and Speech*, **4**: 232-237.

Grice, Martine, Ralf Benzmüller, Michelina Savino, and Bistra Andreeva (1995) "The intonation of queries and checks across languages: Data from Map Task Dialogues," in *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Stockholm, Vol. 3, pp. 648-651.

Grice, Martine and Michelina Savino (1995) "Low Tone versus 'Sag' in Bari Italian Intonation: A Perceptual Experiment," in *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Stockholm, Vol. 4, pp. 658-661.

Grice, Martine, Matthias Reyelt, Ralf Benzmüller, Jörg Mayer, and Anton Batliner (1996) "Consistency in Transcription and Labelling of German Intonation with GToBI," *Proceedings of the International Conference on Spoken Language Processing*, Philadelphia.

Grosz, Barbara and Julia Hirschberg (1992) "Some Intonational Characteristics of Discourse Structure," in *Proceedings of the International Conference on Spoken Language Processing*, Banff, Canada, Vol. 1, pp. 429-432.

Gussenhoven, C. (1984) *On the Grammar and Semantics of Sentence Accents*, Dordrecht: Foris, especially Ch. 3, "The intonation of 'George and Mildred': Post-nuclear generalisations."

Gussenhoven, C. and A.C.M. Rietveld (1991) "An experimental evaluation of two nuclear-tone taxonomies," *Linguistics*, **29**: 423-449.

't Hart, Johan, René Collier, and Antonie Cohen (1990) *A perceptual study of intonation: An experimental-phonetic approach to speech melody*, Cambridge University Press.

Halliday (1970) *A Course in Spoken English: Intonation*, Oxford University Press.

Hieronymus, James L., and Briony J. Williams (1991) "A comparison of the prosody in read speech and directed monologue in British English," from *Proceedings of the ESCA Workshop on the Phonetics and Phonology of Speaking Styles*, Barcelona, Spain, 30 September to 2 October.

Hirschberg, Julia and Barbara Grosz (1992) "Intonational Features of Local and Global Discourse Structure," from *Proceedings of the DARPA Workshop on Spoken Language Systems*, pp. 441-446.

Hirschberg, Julia and Diane Litman (1987) "Now Let's Talk about *Now*: Identifying Cue Phrases Intonationally," in *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, Stanford, California, pp. 163-171.

Hirschberg, Julia and Diane Litman (1991) "Empirical Studies on the Disambiguation of Cue Phrases," *Computational Linguistics*.

Hirschberg, Julia and Janet Pierrehumbert (1986) "The Intonational Structuring of Discourse," *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, New York, pp. 136-144.

Hobbs, Jerry R. (1990) "The Pierrehumbert-Hirschberg Theory of Intonational Meaning Made Simple: Comments on Pierrehumbert and Hirschberg," in *Intentions in Communication*, Cohen, Morgan and Pollack (eds.), MIT Press, pp. 313-323.

Hockey, Beth Ann (1991) "Prosody and the Interpretation of 'okay'," from *Working Notes of the AAAI Fall Symposium: Discourse Structure in Natural Language and Generation*, Monterey, California, November.

Hockey, Beth Ann (1992) "Prosody and the Interpretation of Cue Phrases," *Proceedings of the IRCS Workshop on Prosody in Natural Speech*, IRCS Report No. 92-37, Institute for Research in Cognitive Science, University of Pennsylvania, pp. 71-77.

Houghton, George (1986) *The Production of Language in Dialogue: A Computational Model*, Ph.D. thesis, University of Sussex.

Houghton, George and Stephen Isard (1987) "Why to speak, what to say and how to say it: Modelling language production in discourse," in P. Morris (ed.) *Modelling Cognition* John Wiley, pp. 249-267.

Isard, Stephen (personal communication) University of Edinburgh, UK.

Jefferson, Gail (1973) "A case of precision timing in ordinary conversation: Overlapped tag-positioned address terms in closing sequences," *Semiotica* **9**: 47-96.

Johns-Lewis, Catherine M. (1986) "Prosodic Differentiation of Discourse Modes," in C. Johns-Lewis (ed.) *Intonation in Discourse*, London: Croom Helm, pp. 199-219.

Kingdon, Roger (1958) *The Groundwork of English Intonation*, London: Longmans.

Knowles, Gerry, Briony Williams and Lita Taylor (1996) (eds.) *A corpus of formal British English Speech*, London: Longman.

Kowtko, Jacqueline (1992) "On the intonation of mono- and di-syllabic words within the discourse framework of conversational games," in *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*, pp. 282-284, Delaware.

Kowtko, Jacqueline (1995) "The function of intonation in spontaneous and read dialogue," in *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Stockholm, Sweden, Vol. 2, 286-289.

Kowtko, J.C., S.D. Isard and G.M. Doherty-Sneddon (1992) "Conversational Games within Dialogue," Research Paper HCRC/RP-31, Human Communication Research Centre, University of Edinburgh.

Kreckel, Marga (1981) *Communicative acts and shared knowledge in natural discourse*, London : Academic Press.

Krippendorff, Klaus (1980) *Content Analysis: An Introduction to its Methodology*, London: Sage.

Ladd, D. Robert (1980) *The structure of intonational meaning: Evidence from English*, Indiana University Press.

Ladd, D. Robert (1996) *Intonational Phonology*, Cambridge University Press.

Lee, Kai-Fu (1990) "Context-Dependent Phonetic Hidden Markov Models for Speaker-Independent Continuous Speech Recognition", in Waibel, A. and K-F. Lee (eds.) *Readings in Speech Recognition*, San Mateo, Calif.: Morgan Kaufmann, pp. 347-365

Liberman, Mark and Cynthia McLemore (1992) "The Structure and Intonation of Business Telephone Openings," *The Penn Review of Linguistics* **16**: 68-83.

Lisker, Leigh (personal communication) Haskins Laboratories, New Haven, Connecticut, USA.

Litman, Diane and Julia Hirschberg (1990) "Disambiguating Cue Phrases in Text and Speech," *Proceedings of COLING*, Helsinki, pp. 251-256.

Macafee, Caroline (1983) *Glasgow*, Amsterdam: John Benjamins.

Macaulay, Ronald K.S. (1994) *The Social Art: Language and Its Uses*, Oxford University Press.

Macaulay, Ronald K.S. (1977) *Language, Social Class, and Education: A Glasgow Study*, Edinburgh University Press.

McClure, J.Derrick (1980) "Western Scottish Intonation: A Preliminary Study," in *Melody of Language: Intonation and Prosody*, (eds.) Waugh, Linda R. and C.H. van Schooneveld, Baltimore, Maryland: University Park Press.

McLemore, Cynthia L. (1991) *The Pragmatic Interpretation of English Intonation: Sorority Speech*, Ph.D. dissertation, University of Texas.

Miller, J. and R. Weinert (1995) "The functions of LIKE in discourse," *Journal of Pragmatics* **23**: 365-393.

Nakajima, Shin'ya and James F. Allen (1993) "A Study on Prosody and Discourse Structure in Cooperative Dialogues," *Phonetica* **50**: 197-210.

Newlands, Alison, Anne H. Anderson and Jim Mullin (1996) "Dialogue Structure and Co-Operative Task Performance in two CSCW Environments," In J. Connelly (Ed.) *Linguistic Concepts and Methods in CSCW*, Springer-Verlag.

O'Connor, Joseph D. and Gordon F. Arnold (1973) *Intonation of Colloquial English: A Practical Handbook*, London: Longman.

Pagel, Vincent, Noëlle Carbonell, Yves Laprie, and Jacqueline Vaissière (1995)

"Spotting prosodic boundaries in continuous speech in French," *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Stockholm, Vol. 4, 308-311.

Palmer, Harold (1922) *English intonation, with systematic exercises*, Cambridge: Heffer.

Palmer, Harold E. and F. G. Blandford (1935) *Everyday Sentences in Spoken English*, 5th Edition (revised), Cambridge: Heffer & Sons, Ltd.

Pierrehumbert, Janet (personal communication) Northwestern University, Evanston, Illinois, USA.

Pierrehumbert, Janet B. (1980) *The Phonology and Phonetics of English Intonation*, PhD dissertation, MIT, reproduced by the Indiana University Linguistics Club, 1987.

Pierrehumbert, Janet and Julia Hirschberg (1990) "The Meaning of Intonational Contours in the Interpretation of Discourse," in *Intentions in Communication*, Cohen, Morgan and Pollack (eds.), MIT Press, pp. 271-311.

de Pijper, Jan Roelof (1983) *Modelling British English Intonation*, Dordrecht: Foris.

Pike, Kenneth (1945) "General Characterisation of Intonation," from *The Intonation of American English* University of Michigan Press, pp. 20-41, in *Intonation: Selected Readings*, (ed.) D. Bolinger, Middlesex: Penguin, 1972, pp. 53-82.

Pitrelli, John F., Mary E. Beckman, and Julia Hirschberg (1994) "Evaluation of Prosodic Transcription Labeling Reliability in the ToBI Framework," in *Proceedings of the International Conference on Spoken Language Processing*, Yokohama, Japan, Vol. 1: 123-26.

Power, Richard (1974) *A Computer Model of Conversation*, Ph.D. thesis, University of Edinburgh.

Power (1979) “The organisation of purposeful dialogues,” *Linguistics* **17**: 107-152.

Price, P. J., W. M. Fisher, J. Bernstein, and D. S. Pallett (1988) “The DARPA 1000-Word Resource Management Database for Continuous Speech Recognition,” *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 651-654.

Price, P. J., M. Ostendorf, S. Shattuck-Hufnagel, and C. Fong (1991), “The Use of Prosody in Syntactic Disambiguation,” *Journal of the Acoustical Society of America* **90**: 2956-2970.

Remez, Robert E., Stefanie M. Berns, Jennifer S. Nutter, Jessica M. Lang, Lila Davachi, and Philip E. Rubin (1991) “On the perceptual differentiation of spontaneous and prepared speech,” Presented at the 121st Meeting of the Acoustical Society of America, Baltimore, Maryland, May.

Remez, Robert E., Philip E. Rubin, and Susan A. Ball (1985) “Sentence intonation in spontaneous utterances and fluently spoken text,” Presented at the 109th Meeting of the Acoustical Society of America, Austin, Texas, April.

Sag, Ivan A. and Mark Liberman (1975) “The Intonational Disambiguation of Indirect Speech Acts,” *Papers from the Eleventh Regional Meeting, Chicago Linguistic Society*, Illinois, pp. 487-497.

Sacks, Harvey, Emanuel A. Schegloff and Gail Jefferson (1974) “A simplest systematics for the organisation of turn-taking,” *Language* **50**: 696-735.

Schegloff, Emanual A. and Harvey Sacks (1973) “Opening-up closings’,” *Semiotica* **8**: 289-327.

Schubiger, Maria (1958) *English Intonation: Its Form and Function*, Tübingen: Verlag.

Searle, John R. (1969) *Speech Acts: An Essay in the Philosophy of Language*, Cambridge University Press.

Shelley, Mary W. (1818) *Frankenstein* London: Dent, 1992.

Siegel, Sidney and N. John Castellan (1988) *Nonparametric statistics for the behavioural sciences*, 2nd edition, London: McGraw-Hill.

Silverman, Kim, Mary Beckman, John Pitrelli, Mari Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert, and Julia Hirschberg (1992) "ToBI: A Standard for Labeling English Prosody," in *Proceedings of the International Conference on Spoken Language Processing*, Banff, Canada, pp. 867-870.

Sinclair, John McH. and R.M. Coulthard (1975) *Towards an Analysis of Discourse: The English Used by Teachers and Pupils*, Oxford University Press.

Sluijter A.M.C. and van Heuven V.J. (1995) "Effects of Focus Distribution, Pitch Accent and Lexical Stress on the Temporal Organization of Syllables in Dutch," *Phonetica* **52**: 71-89.

Stubbs, Michael (1983) *Discourse Analysis: The Sociolinguistic Analysis of Natural Language*, Oxford: Blackwell.

Swerts, Marc and Ronald Geluykens (1994) "Prosody as a marker of information flow in spoken discourse," *Language and Speech* **37**: 21-43.

Taylor, Paul A. (1992) *A Phonetic Model of English Intonation*, Ph.D. thesis, University of Edinburgh. Also (1994) "The rise/fall/connection model of intonation," *Speech Communication* **15**: 169-186.

Taylor, Paul and Alan W. Black (1994) "Synthesizing Conversational Intonation from a Linguistically Rich Input," from *Conference Proceedings of the Second ESCA/IEEE Workshop on Speech Synthesis*, Mohonk Mountain House, New Paltz, NY, 12-15 September, pp. 175-178.

Taylor, P. A., H. Shimodaira, S. D. Isard, S. King, and J. Kowtko (1996) "Using Prosodic Information to Constrain Language Modesl for Spoken dialogue," *Proceedings of the International Conference on Spoken Language Processing*, Philadelphia.

Thompson, Sandra A. and William C. Mann (1987) "Rhetorical Structure Theory: A Framework for the Analysis of Texts," *International Pragmatics Association, Papers in Pragmatics* July, Vol. 1, pp. 79-105.

Traum, David R. and Elizabeth R. Hinkelman (1992) "Conversation acts in task-oriented spoken dialogue," Technical Report 425, June, University of Rochester Computer Science Department, New York. Also *Computational Intelligence Special Issue: Computational Approaches to Non-Literal Language*, Vol. 8, No. 3, August.

Veilleux, Nancy, Mari Ostendorf, Patti J. Price, and Stefanie Shattuck-Hufnagel (1990) "Markov Modeling of Prosodic Phrase Structure," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 777-780.

Walker, Marilyn A. (1992) "When Given Information is Accented: Repetition, Paraphrase, and Inference in Dialogue," *Proceedings of the IRCS Workshop on Prosody in Natural Speech*, IRCS Report No. 92-37, Institute for Research in Cognitive Science, University of Pennsylvania, pp. 231-240.

White, S. (1989), "Backchannels across cultures: A study of Americans and Japanese," *Language in Society* **18**: 59-76.

Wilkes-Gibbs, Deanna (1993) "Studying Language Use as Collaboration," in G. Kasper and E. Kellerman (Eds.), *New Directions in Communication Strategy Research*, London: Longman.

Young, S. J. and C. E. Proctor (1989) "The design and implementation of dialogue control in voice operated database inquiry systems," *Computer Speech and Language*, **3**: 329-353.