COMPARING INTONATIONAL FORM WITH DISCOURSE FUNCTION: A STUDY OF SINGLE WORD UTTERANCES*

Jacqueline C. Kowtko†

Human Communication Research Centre
University of Edinburgh, 2 Buccleuch Place
Edinburgh EH8 9LW U.K.

ABSTRACT

Recent attempts to analyze the function of intonation in discourse (both monologue and dialogue) classify the data according to type of intonational tune [4, 7] and make a more or less general characterization of the discourse function associated with utterances containing the particular tunes [8, 5]. The literature shows convincingly that intonation signals boundaries in discourse structure, but lacks a clear specification of discourse function. A suitable discourse taxonomy is needed to fine-tune the relationship between intonation and discourse function. A recent analysis of dialogue [6] provides a framework of conversational games which allows more fine-grained examination of prosodic function. The current paper introduces an intonational analysis of single word utterances based upon such a framework and compares results in progress with previous work on intonation.

1. INTRODUCTION

Recent approaches to the analysis of intonational function within dialogue include an examination of the tunes carried by single-word *cue phrases* (e.g. *now* [4], *okay* [5], and others [7]) across different discourse situations. The literature also includes a more sweeping approach toward classifying phrase-final tunes which presents broadly generalized discourse functions for each of three types of intonational tune: phrase-final *rise*, *level*, and *fall* [8]. Since there is currently no commonly accepted *grammar* of discourse, these studies devise their own relevant discourse categories. Hockey [5, p. 1] reflects upon the problem, with reference to cue phrases. She states that cue phrases

> ...convey information about the structure of a discourse rather than contributing to the semantic content of a sentence. ... Context and prosody are major factors contributing to differences in interpretation among various instances of a cue phrase. In order to investigate the connection between prosodic features and uses of a cue phrase, uses must be identified.

The above is partly a response to Hirschberg and Litman [4, 7] who limit their description to a binary discourse/sentential distinction. Litman and Hirschberg [7] leave the analysis of cue phrase function to the interpretation of various specific discourse approaches and instead focus on validating their prosodic model of cue phrase use [4] with additional data from monologue. The model specifies that a cue phrase in discourse use will occur either alone in a phrase (with unspecified tune) or initially in a larger phrase (deaccented or with a low tone). Thus, Litman and Hirschberg leave open the question of how their prosodic model could further specify discourse function.

McLemore [8] approaches discourse as structured by topics and interruptions. Her data includes announcements given at Texas sorority meetings and conversation between members. She finds that phrase-final tunes indicate certain general functions: *rising* tune *connects*, *level* tune *continues*, and *falling* tune *segments*. Context determines how each of these tunes operates. For instance, phrase-final rise, indicating non-finality or connection, can manifest itself as turn-holding, phrase subordination, or intersentential cohesion. Likewise, the other tunes perform slight variations on the function of *continue* and *segment* according to context.

Hockey [5] admits to settling upon an arbitrary system of discourse classification after attempting to adopt a previous analysis based upon a somewhat similar set of speech data (trying to map discourse categories from conversation at a library reference desk to talk arising from a paperclip task). She focuses on task oriented dialogue and attempts to specify discourse function of the cue phrase *okay*. She presents her results in terms of intonational contours and their corresponding discourse categories, finding that they correlate with McLemore's [8] results: 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. But her

*An earlier version of this paper appears in the Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics, 1992, 282-284.

†A UK Overseas Research Student Award provides partial support. I would like to thank my advisors Stephen Isard and D. Robert Ladd for advice on this work.

categorization of discourse is still weak in that it is not replicable.

Admittedly, there are a limited number of intonational tunes (low rise, high rise, level, fall, etc.). But limitation in intonational tune should not force a limitation in discourse category. Detailed understanding of intonational function is necessarily linked to a more robust view of discourse structure. These previous studies provide good intonational analysis but within vague discourse structures.

2. CONVERSATIONAL GAMES IN DIALOGUE

The analysis offered by Kowtko, Isard, and Doherty-Sneddon [6] provides an independently defined taxonomy of discourse structure which allows a closer examination of how intonation signals speaker intention within task oriented dialogue. In the analysis, linguistic exchanges termed *conversational games* (from a tradition of literature originating in [9]) embody the *initiation–response–feedback* patterns which relate to underlying non-linguistic goals. It is through the framework of games and their components, *conversational moves*, that the intonation of single word utterances can be compared with their discourse function, as intended by the speaker.

A conversational game is defined as consisting of the turns necessary to accomplish a conversational goal or sub-goal. The initiating utterance determines which game is being played and is similar to the *core speech act* in Traum and Allen [10]. In the terms of Clark and Schaefer [3], the initiating utterance serves as *presentation* phase and the ensuing *response* and *feedback* moves function primarily as *acceptance* phases. Implicit, mutually agreed rules dictate the shape of a game and what constitutes an acceptable move within a game. These rules embody procedural knowledge which speakers employ in everyday conversation.

The repertoire of games and moves in Kowtko *et al.* [6] is based upon a map task (see [1], for a detailed description): One person is given a map with a path marked on it and has to tell another person how to draw the path onto a similar map. Neither participant can see the other's map.

The nature of the map task is such that from the conversations the speaker's intentions remain fairly obvious. Kowtko *et al.* [6] report that one expert and three naïve judges agree on an average of 83% of the moves classified in two map task dialogues. Six games appear in the dialogues: Instructing, Checking, Querying-YN, Querying-W, Explaining, and Aligning. They are initiated by the following moves:

INSTRUCT	Provides instruction
CHECK	Elicits confirmation of known information
QUERY-YN	Asks yes-no question for unknown information
QUERY-W	Asks content, *wh-*, question for unknown information
EXPLAIN	Gives unelicited description
ALIGN	Checks alignment of position in task

Six other moves provide response and additional feedback.

CLARIFY	Clarifies or rephrases given information
REPLY-Y	Responds affirmatively
REPLY-N	Responds negatively
REPLY-W	Responds with requested information
ACKNOWLEDGE	Acknowledges and requests continuation
READY	Indicates intention to begin a new game

Since the map task involves one player instructing another on how to draw a path, the conversation naturally consists of many Instructing games. The structure of games allows for looping of response and feedback moves within a game and nesting of games.[1]

The prototypical game consists of two or three moves: initiation, response, and optionally feedback. The large majority of games (84% from a sample of 3 dialogues, $n = 65$) match the simple prototype. Games that do not match the prototype are still well-formed, having extra response–feedback loops, nested games, or extra moves. Very few games (less than 2%) break down as a result of a misunderstanding or other problem.

Here is an example of a prototypical Instructing game. The vertical bar indicates the boundary of a move:

A: Right,|| just draw round it.
READY || INSTRUCT
B: Okay.
ACKNOWLEDGE

Conversational game structure offers a taxonomy which specifies both the function and context of an utterance, as move x within game y. This facilitates the study of the function of intonational tune, since the tune reflects

[1] As a comparison with Clark and Schaefer [3] embedded games often coincide with instances of embedded contributions in the acceptance phase.

an utterance's conversational role.

3. INTONATION IN GAMES

Using data from map task dialogues [1], I have been analyzing single words which compose moves within themselves: *right, okay, aye,*[2] *yes, no, mmhmm,* and *uh-huh.* They typically surface as 5 of the 12 moves in the games analysis [6]: READY, ACKNOWLEDGE, ALIGN, REPLY-Y, and REPLY-N. The current data set consists of 56 out of 80 single word moves spoken by 3 of the 4 conversants in 2 dialogues. For purposes of this study, I am excluding words which form partial utterances (24 of the 80), thus avoiding any interference with accents in the speakers' larger intonational phrases. I have intonationally transcribed each word as high level (H), low level (L), rise (LH), fall (HL), rise-fall (LHL), and fall-rise (HLH).

In order to compare my results with those of McLemore [8] and Hockey [5], I have tried to interpret each utterance as belonging to one of the three general, functional categories. Certain trends become visible: ACKNOWLEDGE moves after EXPLAIN or INSTRUCT, which interrupt the speaker without taking control, typically *connect*; READY and ACKNOWLEDGE moves which precede other moves by the same speaker *continue*; REPLY-Y, REPLY-N, and ACKNOWLEDGE after EXPLAIN or a response move (specifically elicited moves) *segment.*

The data yield the results shown in Table 1.[3]

Table 1: Intonational Tune vs. Dialogue Function

	Connects	*Continues*	*Segments*	
Rising	**1**	0	5	17%
Level	12	**3**	20	9%
Falling	1	0	**14**	93%
	7%	100%	36%	

From the table, we see that 17% of *rises* appear as *connecting* moves, 9% of *levels* as *continuing* moves, and 93% of *falls* as *segmenting* moves. Only the last category matches other published results. Similarly, analyzing the data according to general discourse function (looking down the columns) reveals that only one of the three categories appears to match previous results: *continuing* moves have a *level* intonational tune. It is possible that my classification of utterances would not be corroborated and cause some of the disagreement. Also, it is possible that dialectal variation would account for some of the difference, but I believe that these factors do not account for the difference in results.

These results reflect an intonation-based approach. Information may be lost in the process of collapsing various discourse contexts into three intonational categories (as in [8]) and then limiting discourse categories to match those three existing intonational categories (as in [5]). Using independently motivated discourse categories, in a discourse-based approach, should allow one to see clearer, more detailed results.

When categorized according to *move* (specific function) and position in *game* (discourse context), trends begin to emerge from the data. Granted, the numbers for each category are currently small and not statistically reliable, but some trends are striking and suggest that more data will prove to yield interesting results. Of the 56 data points considered here, three moves are represented: REPLY-Y, REPLY-N, and ACKNOWLEDGE. We find that when one of the utterances appears as a REPLY-Y move in an *Aligning* game, the tune will be *level* if the game is embedded, otherwise *falling.* REPLY-Y and REPLY-N moves within *Querying-YN* games vary according to the previous speaker's last accent. The tune is *low level* when the previous speaker ends low and *falling* when the previous speaker ends high. A single word appearing as an ACKNOWLEDGE in an *Explaining* game generally carries a *low level* tune. When in an *Instructing* game, it carries a *falling or rising* tune after an ALIGN move or continued INSTRUCT move, and otherwise a *level* tune. Within a *Querying-YN* game, there are not yet any clear patterns for the ACKNOWLEDGE move, as half of the tunes are *level* and half *falling.* Within a *Querying-W* game, the tune is *rising.* These results are summarized in Table 2.

I have considered two theories as to why the previous speaker's last accent influences the tune of a conversational move. Firstly, there is the possibility that both speakers cooperatively maintain an overall tune (through *pitch concord* which matches the final key of one move and the initial key of the next move [2]). However, if this were the case, we would expect to see more influence from the previous speaker's accents in other categories of conversational move. More likely is the second possibility that the difference in last accent represents a different nuance of meaning, to which the hearer then responds appropriately. The question of what influences the previous speaker's last accent in a move remains unknown.

[2]Participants in the map task were taken from the population of undergraduates at Glasgow University, and the dialogues consequently contain Scottish English.

[3]The score of 93% is significant ($p < .01$). The 7% is also significant ($p < .01$) and the 9% borderline ($p < .05$), although opposite to predicted results. All other results are statistically non-significant ($p > .05$), according to the Kolmogorov-Smirnov One-sample Test.

Table 2: Intonation Associated with Move X in Game Y

Move	*Game*	*Tune*	*Additional Factors*	*Data*
REPLY-Y	Aligning	H/L	game is embedded	5 of 5
		HL	otherwise	1 of 1
REPLY-Y/N	Querying-YN	L	prev. speaker ends low	5 of 5
		HL	prev. speaker ends high	5 of 6
ACKNOWLEDGE	Explaining	L		3 of 4
ACKNOWLEDGE	Instructing	LH/HL	after ALIGN or continued INSTRUCT (i.e. elicited)	8 of 10
		L/H	otherwise	15 of 18
ACKNOWLEDGE	Querying-YN	L/H	no clear pattern	3 of 3
		HL	no clear pattern	3 of 3
ACKNOWLEDGE	Querying-W	LH		1 of 1

Work is progressing on other dialogues, amassing enough pitch trace data to allow clear patterns to emerge for each type of move in each game context. The goal is, within a discourse context, to be able to predict an utterance's function or *move*, given the intonation, and, conversely, predict intonation, given the type of move.

References

1. Anderson, Anne H., Miles Bader, Ellen G. Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry Thompson, and Regina Weinert, "The HCRC Map Task Corpus," *Language and Speech*, Vol. 34, 1991, No. 4, pp. 351-366.

2. Brazil, David, Malcolm Coulthard and Catherine Johns, *Discourse Intonation and Language Teaching*, Longman, London, 1980.

3. Clark, Herbert H. and Edward F. Schaefer, "Collaborating on contributions to conversations," *Language and Cognitive Processes*, Vol. 2, 1987, No. 1, pp. 19-41.

4. Hirschberg, Julia and Diane Litman, "Now let's talk about *now*: Identifying cue phrases intonationally," *Proceedings of the 25th annual Meeting of the Association for Computational Linguistics*, 1987, pp. 163-171.

5. Hockey, Beth Ann, "Prosody and the interpretation of 'okay'," Presented at the *AAAI Fall Symposium on Discourse Structure in Natural Language Understanding and Generation*, November, 1991.

6. Kowtko, Jacqueline, Stephen Isard and Gwyneth Doherty-Sneddon, "Conversational games within dialogue," Research Paper HCRC/RP-31, Human Communication Research Centre, University of Edinburgh, 1992.

7. Litman, Diane and Julia Hirschberg, "Disambiguating cue phrases in text and speech," COLING-90 *Proceedings*, 1990, pp. 251-256.

8. McLemore, Cynthia A., *The Pragmatic Interpretation of English Intonation: Sorority Speech*, Ph.D. dissertation, University of Texas at Austin, 1991.

9. Power, Richard, *A Computer Model of Conversation*, Ph.D. dissertation, University of Edinburgh, 1974.

10. Traum, David R. and James F. Allen, "Conversation Actions," *Proceedings of the AAAI Fall Symposium on Discourse Structure in Natural Language Understanding and Generation*, November, 1991, pp. 114-119.