

ON THE INTONATION OF MONO- AND DI-SYLLABIC WORDS WITHIN THE DISCOURSE FRAMEWORK OF CONVERSATIONAL GAMES

Jacqueline C. Kowtko*
Human Communication Research Centre
University of Edinburgh
2 Buccleuch Place
Edinburgh EH8 9LW SCOTLAND
Internet: J.Kowtko@edinburgh.ac.uk

Abstract

Recent studies on the analysis of intonational function examine a range of materials from cue phrases in monologue (Litman and Hirschberg, 1990) and dialogue (Hirschberg and Litman, 1987; Hockey, 1991) to longer utterances in both monologue and dialogue (McLemore, 1991). Results match specific intonational tunes to certain discourse functions which are more or less well defined. Although these results make a convincing case that intonation does signal a change in discourse structure, the specification of discourse function remains vague. A suitable taxonomy is needed to fine-tune the relationship between intonation and discourse function. A recent analysis of dialogue (Kowtko et al., 1991) provides a framework of conversational games which allows more fine-grained examination of prosodic function. The current paper introduces an intonational analysis of mono- and di-syllabic words based upon such a framework and compares results in progress with previous work on intonation.

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* (Hirschberg and Litman, 1987), *okay* (Hockey, 1991), and others (Litman and Hirschberg, 1990)) 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* (McLemore, 1991). Since there is currently no workable *grammar* of discourse, these studies devise their own relevant discourse categories. Hockey (1991, 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 (1987; Litman and Hirschberg, 1990) who limit their description to a binary discourse/sentential distinction. Litman and Hirschberg (1990) leave the analysis of cue phrase function to the interpretation of various specific discourse approaches and instead focus on validating their (1987) prosodic model of cue phrase use 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 (1991) 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*. The specifics about how each of these tunes operates depends upon the context. For instance, phrase-final rise which indicates non-finality or connection manifests itself as turn-holding in one context, phrase subordination in another, and intersentential cohesion in yet another context. Likewise, the other tunes perform slight variations on the function of *continue* and *segment* according to context, which is left up to the reader to determine.

Hockey (1991) admits to settling upon an arbitrary discourse classification and letting her data

*A UK Overseas Research Student Award provides partial support. Thanks to my advisors Stephen Isard and D. Robert Ladd for comments on drafts.

speak for itself, after attempting to adopt a system of analysis based upon a somewhat similar set of speech data¹. 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 (1991) results: 89% of *rising* contour occurs where the speaker was *passing* up a turn and letting the other person continue; 86% of *level* contour serves to *continue* an instruction; 88% of *falling* contour marks the *end* of a subtask. But her categorization of discourse is still weak.

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 weak discourse structures.

Conversational Games in Dialogue

The analysis offered by Kowtko, Isard, and Doherty (1991) 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 Power (1974)) 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 mono- and di-syllabic words 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 (1991). The ensuing *response* and *feedback* moves function as *presentation* and *acceptance* phases, in the terms of Clark and Schaefer (1987). Implicit, mutually agreed rules dictate the shape of a game and what constitutes an acceptable move within a game. These rules embody procedural, as opposed to declarative, knowledge which speakers employ in everyday conversation.

¹Hockey had hoped to map discourse categories of *okay* based upon data collected from conversation at a library reference desk to that arising from a task in which one person described a design for another person to make out of paper clips.

The repertoire of games and moves in Kowtko, Isard and Doherty (1991) is based upon a map task (see Anderson et al., 1991, 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, Isard, and Doherty (1991) 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: Instruction, Confirmation, Question-YN, Question-W, Explanation, and Alignment. 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), and 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), and READY (Indicates intention to begin a new game).

Since the map task involves instructing one player on how to draw a path, the conversation naturally consists of many Instruction games. The structure of games allows for nesting of games and looping of response and feedback moves within games²

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 Instruction game. The vertical bar indicates the boundary of a move:

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

²As a comparison with Clark and Schaefer (1987) embedded games often coincide with instances of embedded contributions in the acceptance phase.

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 an utterance's conversational role.

Intonation in Games

Using data from map task dialogues (Anderson et al., 1991), I have been analyzing mono- and disyllabic words which compose single moves within themselves: *right, okay, yes, no, mmhmm*, and *uh-huh*. In addition, I am categorizing the cases where these words form part of a move. They typically surface as 5 of the 12 moves in the games analysis (Kowtko et al., 1991): READY, ACKNOWLEDGE, ALIGN, REPLY-Y, and REPLY-N. The current data set consists of 68 utterances spoken by 3 of the 4 conversants in 2 dialogues.

In order to compare my results with those of McLemore (1991) and Hockey (1991), I have tried to collapse moves and their contexts into the three general categories: ACKNOWLEDGE move following INSTRUCT serves to *connect*; READY, ACKNOWLEDGE (and other) moves which interrupt an INSTRUCT (i.e. precede a continued INSTRUCT move) *continue*; REPLY-Y, REPLY-N, ACKNOWLEDGE after EXPLAIN, and ACKNOWLEDGE after a response move (specifically elicited moves) *segment*.

The data yield the following results³: 42% of *rises* (5 of 11) appear as *connecting* moves, 30% of *levels* (13 of 44) as *continuing* moves, and 69% of *falls* (9 of 13) as *segmenting* moves. Only one category approaches a match to other published results. It is possible that my decisions of which moves collapse together would not be corroborated and cause some of the disagreement. It is also possible that dialectal variation would account for some of the difference (The map task contains Scottish as opposed to American English), but it would be folly to wave such a hand of dismissal. These results reflect an intonation-based approach. Information may be lost in the process of collapsing various discourse contexts into three intonational categories (McLemore, 1991) and then limiting discourse categories to match those three existing intonational categories (Hockey, 1991). Separate discourse categories, in a discourse-based approach, should facilitate clearer results.

When categorized according to *move* and discourse context, the data begins to speak on its

³ $p > .20$ for each result, according to the Kolmogorov-Smirnov One-sample Test, indicates statistical non-significance.

own. 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. For example, of 15 REPLY-Y/N moves, 12, or 80%, are levels, the 3 others being falls in a single category, REPLY-Y after QUERY-YN. All 4 cases of REPLY-Y after ALIGN are high levels, while REPLY-Y/N after QUERY-YN are mostly low levels (6 of 8).

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, given a discourse context, to be able to predict an utterance's function or *move*, given the intonation, and, conversely, predict intonational tune, given the type of move.

References

- 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 (1991). The HCRC Map Task Corpus. *Language and Speech*, 34(4):351-366.
- Clark, Herbert H. and Edward F. Schaefer (1987). Collaborating on contributions to conversations. *Language and Cognitive Processes*, 2(1):19-41.
- Hirschberg, Julia and Diane Litman (1987). Now let's talk about *now*: Identifying cue phrases intonationally. *Proceedings of the 25th annual Meeting of the Association for Computational Linguistics*, Stanford, 163-171.
- Hockey, Beth Ann (1991). Prosody and the interpretation of "okay". Presented at the *AAAI Fall Symposium*, Monterey, CA, November.
- Kowtko, Jacqueline, Stephen Isard and Gwyneth Doherty (1991). Conversational games within dialogue. *Proceedings of the ESPRIT Workshop on Discourse Coherence*, Edinburgh, April. To appear as an HCRC Research Report, Human Communication Research Centre, Edinburgh, 1992.
- Litman, Diane and Julia Hirschberg (1990). Disambiguating cue phrases in text and speech. *COLING-90 Proceedings*, Helsinki, 251-256.
- McLemore, Cynthia A (1991). *The Pragmatic Interpretation of English Intonation: Sorority Speech*. Ph.D. dissertation, University of Texas at Austin.
- Power, Richard (1974). *A Computer Model of Conversation*. Ph.D. dissertation, University of Edinburgh.
- Traum, David R. and James F. Allen (1991). Conversation Actions. *Proceedings of the AAAI Fall Symposium*, Monterey, CA, November, 114-119.