Proceedings of the IRCS Workshop on Prosody in Natural Speech, August 5-12 1992

Prosody and the interpretation of cue phrases

Beth Ann Hockey

Department of Linguistics
University of Pennsylvania
Philadelphia, PA

ABSTRACT

Cue phrases such as *okay* and *uh-huh* are often multiply ambiguous. Native speakers' intuitions are that the various interpretations of these items are distinguished prosodically. Studies by Hirschberg and Litman [1, 2] confirm these intuitions for cue and non-cue uses of several items. This study shows that various cue uses of an item can also be distinguished prosodically. Based on data from task oriented dialogs, three recurring pitch contours were found to correlate with the presence or absence of two features of the discourse: pronominal anaphora and turn taking.

1. Introduction

Certain linguistic expressions, termed 'cue phrases' [3], or 'discourse markers' [4], convey information about the structure of a discourse rather than contributing to the semantic content of a sentence. Since cue phrases overtly mark discourse information they have great potential as a diagnostic for discourse structure. However a property of cue phrases, noted by [3] and [1], is that they are generally ambiguous at least between discourse (cue) and sentential (non-cue) uses, and often among multiple cue uses as well. Hirschberg and Litman[1] and Litman and Hirschberg[2] report that cue and non-cue uses of many items can be distinguished by a combination of intonational phrasing and type of pitch accent.

Native speakers have strong intuitions that cue phrases such as "okay" can have many interpretations and that the various interpretations can be distinguished prosodically [5]. Given these intuitions, it seems likely that prosody can contribute to distinguishing multiple cue uses from each other as well as distinguishing them from non-cue use. If speakers' intuitions on the disambiguating effect of prosody with relation to cue phrases are accurate, one expects to find at worst each prosodic category correlated to a relatively small number of interpretive categories, so that the prosodic information at least narrows the available choices of interpretation.

Cue use interpretations can range over at least semantic, pragmatic, discourse and interactional factors [4] [3] [6] [7] [8] [9] [10] [11] [12]. Rather than hypothesize at the outset about interpretations, I will focus on identifiable features of the context that correlate with the differential distribution of various pitch contours. One of the goals in developing this type of classification technique is to produce theory-independent and relatively objective diagnostics of discourse structure. The descriptive results of such techniques can then be used to investigate the adequacy of a variety of discourse models in relation to actual discourses.

2. Methods

Data for the study is from taped dialogs generated by a task requiring two participants separated by a barrier to cooperatively reconstruct a paperclip design. These conversations are each about twenty minutes long and provide a fairly large number of cue phrases. This paper examines three of these paperclip task conversations with a total of four speakers. Speakers are identified by first initial and number. The conversations are represented in the tables in the next section by a sequence of two speaker identifications. The speaker who started with the completed design is listed before the speaker who was trying to reconstruct the design.

The relation between prosody and interpretation of cue words is investigated by forming natural groupings of F0 contours and by coding for certain properties of contexts, and identifying correlations between the F0 groupings and the context properties.

Grouping of F0 contours was done using characteristics such as relative F0 height of the first and second syllables and general shapes of the two syllables (e.g. rise, fall, level, degree of rise or fall). For each lexical item it was relatively easy to divide tokens into natural intonational classes by sorting pitch contours visually and auditorily, without relying on any previously-assumed system of description. This classificatory independance is an advantage since the fit between existing descriptive systems and natural data is often dubious [13]. Only tokens that were the sole items in some level of intonational phrase were used. This includes all tokens that constitute an entire utterance by themselves and tokens with sufficient phrasal separation so as not to be part of

a larger pitch contour.

Analysis of contexts was done with as few assumptions about discourse structure as possible and without reference to a specific theory of discourse. Two factors were considered in the classification of contexts:

1. distribution of pronominal anaphora and
2. turn taking behavior

As was done in Walker and Whittaker[14], I take the distribution of anaphora to be an indirect indicator of discourse structure. The arguement for using anaphora distribution as an indicator of discourse structure is based on the widespread observation (e.g. [3] [15]) that pronominal anaphora to an antecedent outside the discourse segment(s) containing the pronoun is generally not possible. Therefore, in relation to cue phrases, one expects that if a particular instance of a cue phrase is associated with a discourse segment boundary, a pronoun following that cue phrase should not be able to have an antecedent which preceeds the cue phrase. Each conversation used as data was coded from the transcript for the locations of anaphoric elements and their antecedents. Both pronouns and definite noun phrases were included in the anaphora database. Using the anaphora database, cue phrases could then be examined for whether either pronominal or definite NP anaphora took place across them. Notice that this analysis of anaphora does not take into account any details of discourse structure but only whether any pronoun following a cue phrase has an antecedent anywhere in the discourse preceding the cue phrase.

The turn taking behavior was coded for changes of speaker vs. continuation of a turn. The binary distinction is based on whether the same speaker talked immediately following the item or whether the other speaker talked.

Analysis was first done on G1B1 for the lexical item "okay". The results of that analysis were then tested on B2C1 and C1B2 for "okay" and on all three data sets for the item "uh-huh".

3. Results

3.1. G1B1 "okay"

A striking result from the intonational anaylsis of this first data set is that three especially clear contours emerge from the visual and auditory classification procedure.

One F0 contour type (ct1) is flat. The two syllables have very close F0 values and each syllable remains at its value for most of the syllable duration.

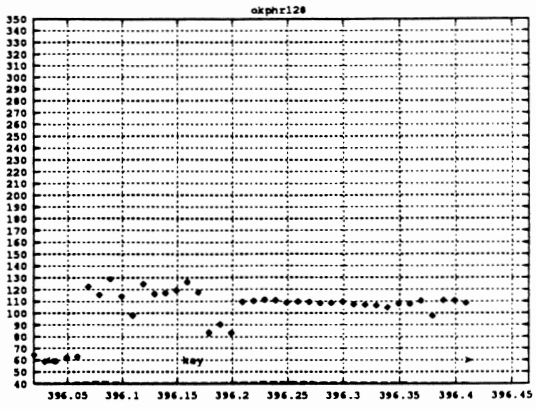

Figure 1: ct1

The second contour type (ct2) has a first syllable higher than the second with an abrupt transition. Both syllables have constant F0 value so are basically flat.

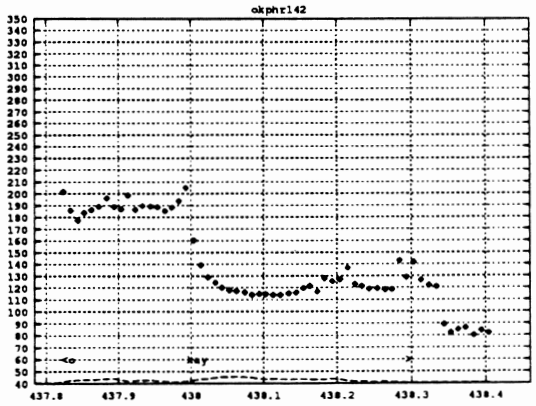

Figure 2: ct2

In the third contour type (ct3), the first syllable is flat or slightly falling. The second syllable is rising. The second syllable begins higher than the end of the first and ends considerably higher than any point in the first syllable.

The results on the first data set show that each of the three most prevalent F0 contours correlates with a distinct context that can be identified by a combination of pronominal anaphora phenomena and and turn taking behavior. One F0 contour type(ct1) is flat. The two syllables have very close F0 values and each syllable remains at its value for most of the syllable duration.

Pronominal anaphora occurs across none of the 8 tokens of ct2, while it does occur in 5 of 13 tokens of ct1 and ct3. This supports the claim in [16] that ct2 was associated

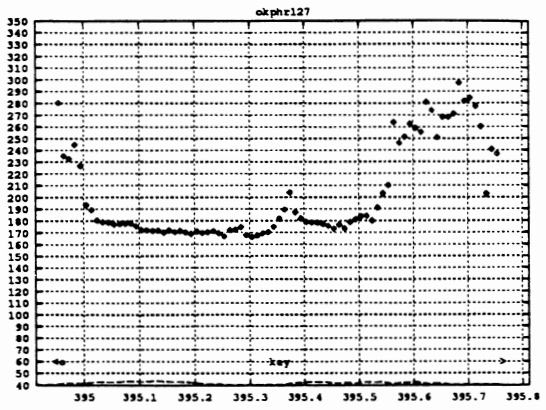

Figure 3: ct3

with the end of a discourse segment. Ct1 is always a turn continuation while ct3 is always a turn change, so ct1 and ct3 can be distinguished on that basis (see table 1 below).

G1B1	pronominal anaphora	turn change
ct1	25%	0%
ct2	0%	0%
ct3	60%	100%

Table 1: Occurance of pronominal anaphora and turn change across contour types in dialog G1B1

3.2. B2C1 and C1B2 – "okay"

The table below shows the distribution of contour types across instances "okay" for all three data sets.

		ct1	ct2	ct3
okay	G1B1	8	8	5
	B2C1	3	10	13
	C1B2	7	4	8

Table 2: Occurance of contour types with *okay* across dialogs

The two tables below show the occurance of pronominal anaphora and turn change for the second and third data sets. The categorical association of ct3 with a turn change observed in G1B1 is also present in B2C1 and C1B2.

B2C1	pronominal anaphora	turn change
ct1	1 (34%)	2 (67%)
ct2	1 (10%)	6 (67%)
ct3	4 (31%)	13 (100%)

Table 3: Occurance of pronominal anaphora and turn change across contour types in dialog B2C1

C1B2	pronominal anaphora	turn change
ct1	1 (34%)	2 (67%)
ct2	2 (50%)	3 (75%)
ct3	1 (13%)	8 (100%)

Table 4: Occurance of pronominal anaphora and turn change across contour types in dialog C1B2

3.3. Uh-huh

Two of the three recurrent contours found with *okay* also occur with *uh-huh*. The minor differences between a contour type occuring on *okay* and the same contour type occuring on *uh-huh* can be attributed to segmental effects. As can be seen in table 5 below ct2 does not occur with *uh-huh*. The number of occurances of ct1 and ct3 varies considerably across speakers.

		ct1	ct2	ct3
uh-huh	G1B1	12	0	40
	B2C1	2	0	12
	C1B2	1	0	10

Table 5: Occurance of contour types with *uh-huh* across dialogs

Only data for turn changes is shown in the table 6 below since the presence or absence of pronominal anaphora only distinguishes ct1 and ct3 from ct2 but not from each other. Pronominal anaphora did occur across both ct1 and ct3 with *uh-huh* as was the case for *okay*.

	turn change
ct1	13 (87%)
ct3	61 (98%)

Table 6: Occurance of turn change across contour types for *uh-huh*

4. Discussion

4.1. Ct3 and turn change

Across four speakers and two different lexical items ct3 categorically marks a turn change. The one instance of ct3 on "uh-huh" which is not listed in the table as turn change is actually a coding dilemma since in this case immediately following the ct3 marked "uh-huh" , both speakers talk simultaneously. Rather than being a counter example, this instance gives direct evidence that the other speaker actually did interpret the "uh-huh" with ct3 as signaling a turn change. The transcript of this instance is shown below, brackets indicate simultaneous talk.

(1)

222 B1: uh-huh. [fat?]

223 G1: [and the] fat end will be facing away from you

4.2. Ct2 and the absence of pronominal anaphora

This data clearly shows that the presence of ct2 on an instance of "okay" proscribes pronominal anaphora across that instance. There are no instances of "uh-huh" with ct2. This is consistent with an analysis of ct2 as marking a discourse segment boundary. Intuitively "uh-huh" cannot be used to mark discourse segment boundaries. We would therefore predict that "uh-huh" would be incompatible with ct2, and this is in fact the case. What particular properties of "uh-huh" are responsible for this difference in distribution between it and "okay" is a topic of ongoing research.

Three instances of ct2 in B2C1 and C1B2 seem to contradict this analysis of ct2 as a boundary marker. However, in examining these 3 instances, it is clear that they are only apparent counter examples to the generalization that ct2 marks the end of a segment.

(2)

25 B2: black white red yellow

26 C1: black white red yellow should these be linked?

27 B2: and these are all linked right

28 C1: okay

29 C1: black white red yellow, okay

30 B2: okay and I would say these put them from the the distance from the top to the bottom should be maybe ten inches

(3)

99 C1: uh and the black one is going to come up just about below just a hair below where the other black one is in the middle

100 B2: okay

101 B2: so it's about on the same level as the blue one in the in the far line

(4)

216 B2: [over an inch to the right] [(())]

217 B2: okay

218 C1: you got that

219 B2: yeah

All three examples are explained if a speaker can only end his or her own discourse segment. So in (2) the speaker ends the segment consisting of his utterance in line 29. Since line 29 is embedded in B2's larger segment, which includes lines 25, 26, 27 and 30, the closure of 29 does not affect the ability of "these" in 30 to have "these" in 27 as an anaphoric antecedent. Similarly, the "okay" on line 217 in (3) has only itself to end, so does not efect the subsequent anaphora to C1's utterance. In (4) the

"okay" can end at most itself and line 216. Notice that in this case C1 seeks additional feedback after B2's "okay" before continuing, since he does not have the clear turn change marking that would have been provided had B2 used ct3 instead.

4.3. Ct1

The clear difference in turn taking behavior between ct1 and ct3 in G1B1 was not observed in the other two data sets. So for the data as a whole ct1 and ct3 cannot be distinguished by the discourse features discussed in this paper in cases where there is a turn change. Ct1 simply conveys nothing to a hearer about whether not a turn change will follow. Obviously there are many discourse features that have not been discussed in this paper that might differentiate ct1 and ct3 in all environments. Research is underway to investigate such additional discourse features. Another question which needs to be addressed in relation to ct1 is why such a striking difference between G1B1 and the other two data sets. The answer to this question may involve issues of individual and sociolinguiguistic variation. Another possibility is that there is a difference related to percieved need for clarity by the speakers. If speakers have to make extensive corrections of prior discourse perhaps the frequency of the ambiguous ct1 would be reduced in the corrective dialog. These issues need to be addressed by future research on additional discourse features and on additional data.

4.4. Prosody and interpretation

The data shows how two of the contours discussed in this paper reduce the range of interpretation a hearer can have for the cue phrases with which the contours are associated. In working toward results that can explain how a hearer can use prosody, it is valuble to start from a data driven analysis of the prosody. The prosody is part of the speech signal in a way that abstract discourse categories are not. We would like to have an account which explains how the prosody provides information about interpreting the discourse and not the other way around. It is unlikely that a top-down approach to prosody would have led to investigating the particular contours discussed in this paper. It seems even less likely that top-down approach based on discourse categories would.

In addition to showing how prosody can reduce the range of interpretation of a cue phrase, these results suggest that in some uses of cue phrases the prosody matters more than the particular lexical item. For example, "Okay" and "uh-huh" function in the same way when uttered with ct3. Although the number of analyzable tokens of other cue phrases in this data is too small to arrive at any conclusions, analysis of the usable tokens suggests that a number of other items, such as "right", "yeah", "alright" and "so", can serve the same function as "okay" and "uh-huh" when uttered with ct3. It also appears that like "okay", "right" and "alright" with ct2 can function as discourse segment boundaries. Cue phrases are not entirely vacuous semantically and the semantics of the item can interact to generate implicatures when the semantic content of the item is not being used directly. For example, if "so" is used with ct3, the function is the same as with other items such as "okay" and "uh-huh" namely to prompt the other speaker to talk, to pass up a turn. But with "so" the semantics of "so" which is to conjoin a fact, action or event with its result [4] comes through as an implicature that what should follow should be a result of the prior turn. The person who utters "so" with ct3 when they could have used "okay" or "uh-huh" seems to be conveying that the other person should go on AND get to the point(=result). Ct3 can be thought of as marking the item to which it attaches for an interactional interpretation.

References

1. Julia Hirschberg and Diane Litman. Now let's talk about *now*: Identifying cue phrases intonationally. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, 1987.
2. Diane Litman and Julia Hirshberg. Disanbiguating cue phrases in text and speech. In *Proceedings of Coling90*, 1990.
3. Barbara J. Grosz and Candace L. Sidner. Attention, intentions, and the stucture of discourse. *Computational Linguistics*, 12(3):175–204, 1986.
4. Deborah Schiffrin. *Discourse markers.* Cambridge University Press, 1987.
5. Beth A. Hockey. An experimental approach to invesitgating the role of prosody in the interpretation of cue phrases. Draft, 1992.
6. Lawrence C Schourup. *Common discourse particles in English conversation.* PhD thesis, Ohio State University, 1982.
7. E. A. Schegloff. Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In *Analyzing Discourse: Text and Talk.* Georgetown University Press, 1982.
8. Robin Cohen. A computational theory of the function of clue words in argument understanding. In *Proceedings of Coling84*, 1984.
9. Robin Cohen. Analyzing the stucture of argumentative discourse. *Computational Linguistics*, 13:11–25, 1987.
10. Steve Whittaker and Phil Stenton. Cues and controls in expert-client dialogues. In *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, 1988.
11. Rachel Reichman. *Getting computers to talk like you and me: discourse context, focus, and semantics.* MIT Press, 1985.

12. M. Merritt. On the use of 'okay' in service encounters. In J. Baugh and J. Sherzer, editors, *Language in use*, pages 139–47. Prentice-Hall, Englewood Cliffs, NJ, 1984.

13. Cynthia Ann McLemore. *The Pragmatic Interpretation of English Intonation: Sorority Speech.* PhD thesis, University of Texas at Austin, 1991.

14. Marilyn A. Walker and Steve Whittaker. Mixed initiative in dialogue: An investigation into discourse segmentation. In *Proc. 28th Annual Meeting of the ACL*, 1990.

15. Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. Towards a computational theory of discourse interpretation. Draft, 1986.

16. Beth A. Hockey. Presented at the AAI Fall Symposium, Asilomar CA, November 1991.

okay	talked over	part of larger phrase	**entire utterance**	**own phrase**	total	**percent used**
G1B1	15	44	22	33	114	48%
B2C1	41	42	42	46	170	49%
C1B2	14	10	30	11	65	63%

Table 7: Frequency and phrasing of *okay*

uh-huh	talked over	part of larger phrase	**entire utterance**	**own phrase**	total	**percent used**
G1B1	13	2	60	8	83	82%
B2C1	8	0	25	1	34	76%
C1B2	3	0	13	0	16	81%

Table 8: Frequency and phrasing of *uh-huh*