Possible Lexical Cues for Backchannel Responses

Nigel G. Ward

Department of Computer Science, University of Texas at El Paso, El Paso, Texas, United States
nigelward@acm.org

Abstract

Looking for words that might cue backchannel feedback, I did a statistical analysis of the interlocutors' words preceding 3363 instances of *uh-huh* in the Switchboard corpus. No clear cueing words were found, but collateral findings include the existence of semantic classes that slightly increase the likelihood of an upcoming *uh-huh*, the fact that different classes have their effects at different time lags, and the existence of words which strongly counter-indicate a subsequent *uh-huh*.

Index Terms: uh-huh, feedback, elicitors, temporal distributional analysis, dialog dynamics

1. Background

Regarding the question of when and why listeners backchannel, members of the general public often think that they respond to specific cue phrases. Certainly responses can be elicited, for example by ending any statement with *you know what I mean?*, but the resulting responses are scarcely optional, and thus not strictly backchannels. Today for true backchannels the research on cues focuses elsewhere, namely on non-verbal features, notably features of prosody and gaze [1].

But perhaps lexical cues also do have a role. In the literature there is one relevant suggestion, Allwood's "extremely tentative" identification of the apparent elicitors of feedback in several languages, including for English the words *eh* and *right* [2]; but it seems that no one has ever followed up on this. The question of the existence of lexical cues to backchannels is also of more general interest, as it relates to larger issues regarding the extent of automaticity and responsiveness in dialog.

This paper presents an exploratory study, looking for words that may cue backchannels.

2. Method

Specifically, this paper examines the contexts preceding 3363 occurrences of the most typical backchannel token, *uh-huh*, found in a 650K word subset of Switchboard, a corpus of unstructured two-party telephone conversations among strangers [3]. *Uh-huh* was chosen as the most typical backchannel [4] and also as a word which is almost invariably a backchannel. The method was Temporal Distributional Analysis [5]: this section summarizes this technique and its use for *uh-huh*.

A frequently operative cue word should, by definition, occur frequently in the speech of the interlocutor just before the *uh-huh*. Thus I compiled statistics on the words which commonly preceded *uh-huh*. Lacking foreknowledge of where exactly cues might occur, I compiled statistics at various offsets, as measured from the onset of the context word to the onset of the *uh-huh*. For convenience the offsets were discretized into buckets, thus for example an occurrence of the word *know* starting 1.8 seconds before a *uh-huh* was counted in the 1–2 second bucket.

From the counts over the whole corpus, I computed the degree to which each context word x is characteristic of each bucket t. In particular, I did this by comparing the in-bucket probability to the overall (unigram) probability for x. For example, we can compute the ratio of the probability of *know* appearing in the 1–2 second bucket to the probability of *know* appearing anywhere in the corpus. This we call the R ratio. Specifically, the probability of each word in each bucket, the "bucket probability," is given by its count in the bucket for t divided by the total in that bucket,

$$P_{tb}(w_i@t) = \frac{count(w_i@t)}{\sum_j count(w_j@t)} \tag{1}$$

We can then compute the ratio of this to the standard unigram probability:

$$R(w_i@t) = \frac{P_{tb}(w_i@t)}{P_{unigram}(w_i)} \tag{2}$$

If R is 1.0 there is no connection and no mutual information; larger values of R indicate positive cooccurrence relations, and lower values of R indicate words that are rare in a given context position. To test whether a R-ratio is significantly different from 1.0, I apply the chi-square test, where the null hypothesis is that the context word occurs in a certain bucket as often as expected from the unigram probability of the word and the total number of words in that bucket, where the sample population is relative to all occurrences of *uh-huh* in the corpus.

As my purpose was exploratory, I didn't want to be overwhelmed with possibilities, so I limited attention to the most frequent 5000 words in the corpus, with less frequent words counted as belonging to the out-of-vocabulary class. I also limited attention to words whose R-ratio was significantly high or low, with $p < .001$. Due to the large number of words examined, this does not guarantee that all candidate cues identified are truly significant, and thus the findings are only tentative.

3. Observations

Table 1 shows the result. Within each cell, the words are ordered by extremeness of the R values: above the double line most frequent first, below it least frequent first. I make seven observations:

1. The word *right* is a *counter*-indication to a subsequent *uh-huh*, contrary to Allwood's conjecture, and so are the component words of the phrase *you know*, again contrary to common expectation.

2. Indeed, all discourse markers counterindicate *uh-huh*, except for *um* 2–8 seconds earlier.

3. No word is a truly strong backchannel cue; the highest R-ratio is 3.7 for *time* just before a *uh-huh*. (Note that, by Bayes law, the implications go both ways: since *time* is 3.7 times more common just before *uh-huh*, that means that an observation of the word *time* implies that a backchannel is 3.7 times more likely than usual to occur within a half second.)

4. However there are words which very strongly indicate that a backchannel will *not* occur soon. For example, a *mm-hm* by the interlocutor reduces the likelihood of a backchannel 1–2 seconds later by a factor of 53.

5. Some words do appear somewhat more frequently before *uh-huh*, and these mostly fall in a few categories:

   - The deictics *here*, *there* and *now* are common in the half-second before *uh-huh*.

   - Some pronouns are common at different offsets: *them*, *one*, and *it* just before *uh-huh*; and *I*, *we*, *he*, and *she* 2–8 seconds before.

   - Some verbs, notably *take*, *took*, *went*, *watch*, *use*, *made*, and *had*, are common 1–6 seconds before.

   - Some prepositions are common at various different offsets.

   - Number words are common before *uh-huh* starting about 6 seconds before, as are other expressions of quantity such as *some, little, bit, much* and *more*.

   - Temporal expressions are common at different offsets: *now* just before, and *ago*, *years*, *old*, and *times* 4–8 seconds before.

6. The specific offset makes a difference. For example, the words *I*, *we* and *a* are positive indicators of an upcoming backchannel after certain delays, but they counter-indicate a backchannel right away.

7. The counter-indicating words ("anti-cues") also fall into a few common classes:

   - listenership indicators such as *mm-hum*, and *uh-huh*, unsurprisingly

   - out of vocabulary words ([OOV] in the table), that is, the less frequent words, which include most names, somewhat surprisingly given that some accounts associate *uh-huh* with grounding of new referents

   - cues to starting something new (*that's, it's, I think, yeah, well, oh, uh, um, a* and *the*); after these *uh-huh* is inhibited for a second or so.

4. Discussion

To consider a word to be a cue, it should meet two criteria: it should strongly evoke the backchannel response, and it should have a direct causal effect.

Regarding the first criterion, the ratios seen are not particularly high, especially compared to the strength of gaze and prosodic cues [1], so none of the words identified can be considered a strong cue, if indeed a cue at all.

The second criterion is harder to apply. Determining definitively whether one or more of these words has causal efficacy would require detailed analysis and perhaps controlled experiments. However, a quick examination of a dozen instances of the strongest candidate for cue word status, *time*, when it preceded a *uh-huh*, was not promising: these mostly occurred as part of telling a narrative about some past event, as in *he took a car battery one time* ... and in *one time we used it to pay our rent*. The word *time* in these cases never seemed to bear any special discourse function. More generally, looking over the words above the line in the table, none seems likely to be a cue. One would expect cues to be discourse markers and/or phonetically distinct so that they could be quickly processed and responded to, but these words look like just normal words, bearing their normal meanings and doing their normal functions.

Thus there appears to be no reason to think that that these words are really cues for *uh-huh*. By extension it seems unlikely that there are specific lexical cues for backchannels.

This interpretation, however, raises an interesting question at a deeper level, that of the pragmatic and semantic events that can cue *uh-huh*. The fact that particular words correlate with subsequent backchannels (and others anti-correlate) provides us with clues to what those pragmatic and semantic events might be. Future research

R	8–6 seconds	6–4 seconds	4–2 seconds	2–1 seconds	1–.5 seconds	.5–0 seconds
> 2.8				three, try, used	two	time, here, them
> 2.0	place, old, years	took, went, used, ago, came, started, he's, thought, I'll, home, made, four, far, we've, times	husband, college, take, watch, went	I'll, five, through, never, use, had, school, he's	very, from, real, at, them, in, on, out, into, pretty, their	work, there, now, too, up
> 1.4	she, last, went, family, he's, year, never, here, because, over, my, little, our, those, um, we, were, was, all, mean, would, had	use, she, three, better, my, come, never, other, into, years, only, because, little, um, time, was, one, really, at, we, then, like, I	I'd, my, enough, only, I've, our, maybe, bit, actually, put, she, two, work, been, because, then, little, when, one, we, from, out, he, very, like, was, something, had, get, them, some, I'm, um	doing, around, go, her, these, little, two, an, like, into, work, get, they're, from, as, a, more, been, have, to, up, with, for, on, my, we	little, for, an, much, my, a, your, of, the	one, out, or, it
$< .71$	uh-huh, as	right, oh	oh, right, yeah	that's	I, that's, so, uh, know, you	was, a, I, [OOV], the
$< .50$			[laughter], okay	oh	but, think, [OOV], don't	it's, just, [laughter], have
$< .35$	[OOV], mm-hm	[OOV], mm-hm, uh-huh	[OOV]	okay, yeah, [OOV], [laughter]	if, mean, oh	yeah, we, well, think, they
$< .25$			mm-hm, uh-huh		yeah, [laughter], well	oh, uh, um
$< .18$						if
$< .12$				uh-huh	[vocalized-noise], um	got, would, uh-huh, because
$< .09$					uh-huh	that's
$< .06$						mm-hm
$< .04$					mm-hm	
$< .03$						
$< .02$				mm-hm		

Figure 1: Interlocutor words that are notably frequent and infrequent, as judged by R-ratios, in six regions of time before *uh-huh*.

should look for them.

Doing so would be interesting in many ways, including the fact that, given the varying offsets, this might reveal something about the time constants of the mental processing required to digest information of various kinds and decide to produce a minimal response. That is, these observations could provide an entry to the study of the semantic aspects of dialog dynamics [6, 7], complementing existing work on prosodically- and gaze-cued response patterns.

5. Applications

Quite apart from the possible broader implications suggested above, our results may have practical value as-is.

On the one hand, for the sake of improving the performance of dialog systems that show attention and generate rapport by backchanneling, the findings above, especially regarding the counter-indicators to backchanneling, could be useful.

On the other hand, to improve systems which elicit backchannels using various cues [8], and those which interpret backchannels based on the details of their timing [9], the patterns of co-occurrence could again be useful.

6. Summary

In this study I explored which words tended to precede an *uh-huh* by the interlocutor, using a new statistical analysis method. The results cast doubt on the existence of lexical cues to backchannels, however they do reveal tendencies that suggest new hypotheses about the dynamics of interaction in dialog.

7. References

[1] L.-P. Morency, I. de Kok, and J. Gratch, “A probabilistic multimodal approach for predicting listener backchannels,” *Autonomous Agents and Multi-Agent Systems*, vol. 20, pp. 70–84, 2010.

[2] J. Allwood, “Feedback in second language acquisition,” in *Adult Language Acquisition: Cross Linguistic Perspectives, II: The Results* (C. Perdue, ed.), pp. 196–235, Cambridge University Press, 1993.

[3] ISIP, “Manually corrected Switchboard word alignments.” Mississippi State University. Retrieved 2007 from http://www.ece.msstate.edu/research/isip/projects/switchboard/, 2003.

[4] 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* (D. Tannen, ed.), pp. 71–93, Georgetown University Press, 1982.

[5] N. G. Ward, “Temporal distributional analysis,” in *SemDial*, 2011.

[6] L.-P. Morency, “Modeling human communication dynamics,” *IEEE Signal Processing Magazine*, vol. 27, 2010.

[7] N. G. Ward, “The challenge of modeling dialog dynamics,” in *Workshop on Modeling Human Communication Dynamics, at Neural Information Processing Systems*, 2010.

[8] T. Misu, E. Mizukami, Y. Shiga, S. Kawamoto, H. Kawai, and S. Nakamura, “Toward construction of spoken dialogue system that evokes users’ spontaneous backchannels,” in *Proceedings of the SIGDIAL 2011 Conference*, pp. 259–265, 2011.

[9] T. Kawahara, M. Toyokura, T. Misu, and C. Hori, “Detection of feeling through back-channels in spoken dialogue,” in *Interspeech*, 2008.