

# Can You Predict Responses to Yes/no Questions? Yes, No, and Stuff

Beth Ann Hockey<sup>†</sup> Deborah Rossen-Knill<sup>‡</sup> Beverly Spejewski<sup>†</sup> Matthew Stone<sup>†</sup> Stephen Isard<sup>§</sup>

<sup>†</sup> University of Pennsylvania    <sup>‡</sup> Phila. College of Textiles & Science    <sup>§</sup> University of Edinburgh  
Philadelphia PA 19104-6228    Philadelphia PA 19144-5497    Edinburgh EH1 1HN

{beth, drossen, spejewsk, matthew}@linc.cis.upenn.edu    stepheni@cstr.ed.ac.uk

## Abstract

We analyze what functions as a YES response and a NO response for different yes/no questions. This problem is surprisingly complex: respondents do not always produce overt *yes* or *no* lexical items in response to a yes/no question. In addition, when respondents don't include a clear *yes* or *no* word, they may mean to communicate a clear YES or NO meaning, or something else. We find that the classification of yes/no questions described in (Carletta et al., 1995) for the Edinburgh map task corpus correlates well with whether a response will be a bare *yes* or *no*, a *yes* or *no* plus additional speech, or just speech without an overt *yes* or *no*. Correlation with responses described simply as "direct" or "indirect" is less good. We also find that, under the three-way categorization, the strength of a question's expectation for a YES response predicts the form of the response.

## 1 Introduction

In order to model dialogue, it is necessary to understand what kinds of utterances are likely to occur in a given situation. We must, for example, be able to distinguish different kinds of responses to various types of yes/no questions. While it may at first seem fairly straightforward that the answer to a yes-no question should be *yes* or *no*, or some variant of these terms, it has often been pointed out that acceptable responses may not contain such a term (e.g., (Stenström, 1984; Green and Carberry, 1992; Green and Carberry, 1994; Stubbs, 1983; Ginzburg, 1995). Green and Carberry's (1994) example illustrates this fact:

- (1) a Q: I need a ride to the mall.  
b Are you going shopping tonight?  
c R: [no]  
d My car's not running.  
e The timing belt is broken.

The square brackets indicate that R does not utter *no*, although the response contains a negative answer to the question.

Example (1) raises several questions. When is it appropriate to produce answers that don't contain an overt *yes* or *no* form, but only other speech? When is it sufficient to produce just an overt *yes* or *no*, and when is it appropriate to produce both an overt polar term and other relevant speech? In this paper, we address these questions, based on analysis of natural language data from the Edinburgh map task corpus (Anderson et al., 1991). We consider the whole of a response in relation to its question as communicating a YES, NO or NON-COMMITTAL meaning, and separate the response into one of three categories of form: the bare affirmative or negative terms, which we refer to as *bare yes/no*; polar terms plus additional speech, which we refer to as *yes/no+stuff*; and speech without overt polar terms, referred to as *stuff*.

On analysis, the three-way categorization of form leads to two major observations. First, the form of answer correlates with the function of the question in dialogue as given by the MOVE TYPE of the question (Carletta et al., 1995). Second, the *yes* term used to answer a question correlates with the overall likelihood of producing a YES meaning response for questions of that type—a quality we refer to as the YES-expectation.

## 2 Description of Dialogues Studied

We analyzed eight dialogues averaging seven minutes each of Glaswegian English from the Edinburgh map task corpus ((Anderson et al., 1991), files eaqlc1-eaqlc8). In these dialogues, one person described a route marked on a map, in order for the other person to mark the same route on her own map. Each dialogue had been previously coded for dialogue structure using the system described by Carletta et al. (1995). Their coding scheme has three levels of structure. One of these levels is the conversational game, which consists of an initiating utterance and all utterances following it until the purpose of the game is fulfilled. An initiating utterance and its responses are each classified as particular types of CONVERSATIONAL MOVES based on their purposes. The initi-

Align:  
 “Right?”  
 “Okay? Can you do that?”

Check:  
 “So you don’t have a graveyard?”  
 “Walk below the springs?”

Query-yes/no:  
 “Do you have a fenced meadow?”  
 “Are we going to go below the picket fence?”

Figure 1: Examples of Moves

Bare *yes* (check):  
 Q: And you loop round about a bakery?  
 R: Yeah.

*yes+stuff* (query-yes/no):  
 Q: Have you got a white mountain?  
 R: Yes, I have a white mountain to my north.

*stuff* (query-yes/no):  
 Q: Do you meet the savannah?  
 R: Ehm, when I go round the collapsed shelter my savannah is to my right.

Figure 2: Answer forms

ating utterances of interest for our study are *yes/no* questions. In this study we restricted our attention to questions without negation; negated questions introduce a further level of complexity, which could be better investigated once we have an understanding of the more straightforward cases.

Carletta et al. (1995) classify *yes/no* questions into three different move types: ALIGN, CHECK and QUERY-YES/NO. An align “checks the attention or agreement of the partner, or his readiness for the next move”; a check “requests the partner to confirm information that the checker has some reason to believe, but is not entirely sure about”, and a query-yes/no is any other question that takes an answer of *yes* or *no* (Carletta et al., 1995). Examples of each move type are given in figure 1.

In addition to the coding of move types done by the Edinburgh group, we coded for other features of the answers. We indicated whether or not an answer contained any overt *yes* or *no* terms, which particular *yes* or *no* lexical items it contained, and whether or not it included any other material (see figure 2). We also coded for whether the answer as a whole conveyed a YES or a NO, or a meaning that does not appear to commit to either.

### 3 Results and Analysis

In Green and Carberry (1992; 1994), responses to questions have been broken down into direct and indirect components and studied as such. They cat-

egorize any answer containing an overt polar term as direct, and any other answer which commits to a YES or NO meaning, such as (1)(d–e), as indirect. The categorization reflects a traditional view of responses as direct or indirect speech acts (Searle, 1975).

Using the answers to *yes/no* questions found in these dialogues, we argue that rather than the kind of two-way categorization of answers found in Green and Carberry, it is useful to divide answers into three groups. Rather than categorizing answers on just one dimension, whether or not they contain an overt *yes-no* term, we categorize on two dimensions, whether they contain an overt term, and whether they contain other speech. This elaboration will allow us to make different predictions for the answers to the three different move types, as well as explaining the answers that contain more than a bare *yes* or *no*.

Perhaps what demonstrates the usefulness of the three-way split most clearly is a comparison of two tables classifying response types. The first table results from our initial categorization of responses into direct and indirect categories. In figure 3, we separate YES, NO and NON-COMMITTAL answers, showing how many of each kind of meaning are given as direct and how many as indirect, for each of the kinds of question types we are considering. (Non-committal answers will all be indirect, and these are given as percent of total answers.) This breakdown shows a clear difference between aligns and the other two categories, but the differences between checks and query-yes/no’s are more subtle. Aligns are not answered negatively at all, while the checks and query-yes/no’s both have a large number of direct NO’s plus some indirect NO’s. Aligns also do not have any indirect YES answers, while the other two do.

In figure 4, we have teased apart the direct category into those that consist of just a polar term, and those that also include *stuff*<sup>1</sup>. Now we see a clearer difference between checks and query-yes/no’s in the way that NO answers are formed. Checks don’t have bare NO answers; rather they have a high percentage of NO answers conveyed through an overt polar term and additional speech, and show some possibility for a NO communicated only through *stuff*. Query-yes/no’s have responses of all types.

We can explain these results by appealing to the purpose of each move type and the strength of its YES-expectation. The purpose of an align is to check that the conversation is proceeding smoothly. This offers the hearer the chance to either indicate that all is well, as the normal case, or to point out that the conversation is in need of some repair that the speaker wasn’t aware of. All of the non-committal

<sup>1</sup>The *stuff* typically occurred after the polar term, but not always.

	NO: % of NO's		YES: % of YES's		NON-COMMITTAL % of total
	direct	indirect	direct	indirect	
align	–(0)	–(0)	100%(40)	0%(0)	11%(5)
check	88%(7)	13%(1)	81%(46)	19%(11)	8%(6)
q-y/n	77%(17)	23%(5)	70%(28)	30%(12)	15%(11)

Figure 3: Breakdowns of responses into direct and indirect

	NO			YES			NON-COMMITTAL
	bare	<i>no+stuff</i>	<i>stuff</i>	bare	<i>yes+stuff</i>	<i>stuff</i>	
align	– (0)	– (0)	– (0)	90% (36)	10% (4)	0% (0)	11% (5)
check	0% (0)	88% (7)	13% (1)	37% (21)	44% (25)	19% (11)	8% (6)
q-y/n	50% (11)	27% (6)	23% (5)	35% (14)	35% (14)	30% (12)	15% (11)

Figure 4: Breakdowns of responses into three content forms

answers for aligns in our data serve to initiate a repair, by requesting clarifying information. Although there is some chance of a NO response to an align, the speaker expects a YES—a strong YES-expectation—or else that the hearer will attempt to initiate a repair sequence. As we see in figures 3 and 4, when a YES answer is not valid, the strong YES-expectation of the question renders the respondent unwilling to say NO.

In performing a check, a speaker requests confirmation of information he has some evidence for. This definition implies a greater expectation that the answer will be YES than NO, although not as strongly as with aligns. (Recall that we are only considering non-negated questions.) However, the possibility of NO is clearly not excluded, given that the speaker has some doubt about the information. It seems, then, that we lack bare NO responses because speakers apparently feel compelled to provide information about the unexpected NO. That there are fewer non-committal responses for checks than aligns likely occurs because speakers can produce NO responses, and are therefore not dependent on non-committals to communicate NO.

Finally, query-yes/no's do not have an intrinsic expectation of a YES or a NO—although there may be such expectations for individual occurrences, due to context (Stubbs, 1983). For this move type, we see the greatest number of bare NO's because in this case NO is as informative, cooperative (Grice, 1989) and perhaps polite (Brown and Levinson, 1987) as YES. Put differently, the query-yes/no move does not require *stuff* in a negative response in the way an align and a check do. The difference in YES-expectation also predicts differences in YES answers across move types, as discussed in an expanded version of this paper.

As a final argument for this distinction in expectations, we present in figure 5 the distribution of answers from our data according to their YES/NO mean-

	NO	YES	NON-COMMITTAL
align	0%	89%	11%
check	11%	80%	8%
q-y/n	30%	55%	15%

Figure 5: % of answers having each meaning type

ing, for the different kinds of questions. For aligns, which have a strong YES-expectation, the data show a strong bias toward producing responses that mean YES. The only answers other than a YES do not make a commitment to YES or NO. Checks have a lesser, but still apparent, YES-expectation, and here we see a small number of NO answers. Query-yes/no's, which don't seem to carry any YES-expectation, have plenty of NO answers. What we find through analysis of figure 5, then, is that these built-in expectations, or lack thereof, are particularly useful in accounting for the distribution of content and form in answers.

We also found that the YES-expectation predicts the type of affirmative word chosen in the response. Affirmatives in our data included *yes*, *yeah*, *uh-huh*, *right*, *okay*, *mm-hmm*, and combinations of these. In answers meaning NO, only *no* appeared as an overt polar term in this data. The affirmatives have varying degrees of semantic/pragmatic content. *Yes*, for example, involves a stronger affirmation than *okay* does. The stronger the YES-expectation of the question is, the weaker are the affirmative forms that tend to answer it, as figure 6 demonstrates<sup>2</sup>. For each kind of move, the histogram of *yes* terms falls off fairly smoothly from its peak, but each move has its peak in a different place. Aligns, which have a strong YES-expectation, have a high likelihood of being answered with weak terms such as *okay* and *right*; checks have a somewhat weaker YES-expectation, and tend to be answered by somewhat stronger terms, such as *uh-huh* or *yeah*; and query-yes/no's generally have no

<sup>2</sup>Sometimes more than one overt lexical was used. In those cases, we only counted the first one in figure 6.

	align	check	query-yn
okay	10	1	0
right	21	11	2
mmhmm	7	9	5
uh-huh	7	15	7
yeah	3	15	5
yes	0	1	11
no	0	7	17

Figure 6: number of occurrences of each form

YES-expectation, and the answers to these questions cluster closer to *yes*. Another way of looking at this is to see that the main use of *okay* or *right*, for example, is to answer an align; *uh-huh*, a check; and *yes*, a query-yes/no.

#### 4 Conclusion

Understanding what kinds of answers are to be expected for specific kinds of yes-no questions is an important kind of information that should be available to systems for modeling or interpreting dialogue. Our results and analysis contribute to the understanding of this domain in several ways: providing a new three-way classification of answers; showing that the YES-expectation of a question predicts the class of the response; and showing that the YES-expectation predicts the lexical form of the response.

We find that a number of different kinds of information contribute to the formation of responses to yes-no questions. These include what type of question was asked (and therefore what the speaker's expectations are about the answer), Gricean pragmatic considerations, politeness, and lexical knowledge about the strength of various *yes*-terms. We assume that there are also a number of other contextual factors that were not detailed here. For instance, we expect that a person answering a question will have to weigh the relative importance of politeness versus the need for the other person to be clear what the answer is, and that there may be differences based on the situation or the question in whether explanation of an answer is appropriate. We also expect that the nature of the task the speakers are involved in may affect some dimensions of question asking and answering. We have tried to show that even without including all the possible considerations a speaker may have in formulating an answer to a yes-no question, there are some fairly strong patterns in answers, deriving just from the kind of question that was asked, and our analysis explains these while still allowing some variability based on individuals or situations.

Our results are useful for generation in dialogue systems and for speech recognition. Our results suggest a simple strategy based on the intended answer

and the move type of the user's question for selecting what polar lexical items to use and what kind of *stuff*, if any, to produce in a response. It would be straightforward to follow this strategy in the generation component of a dialogue system because first, move types of the users' questions can be identified using prosody (Taylor et al., 1996), second, that many dialogue systems already incorporate similar kinds of decision networks for choosing forms based on contextual considerations (e.g. (Bateman, 1995; O'Donnell, 1996)).

In speech recognition systems, having tighter constraints on the language model produces better performance. Being able to predict the form of answers from the move type of question they answer provides a valuable source of additional constraints that can be incorporated into such models.

#### References

- A.H. Anderson, M. Bader, E.G. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson, and R. Weinert. 1991. The HCRC map task data. *Language and Speech*, 34(4):351–366.
- J. A. Bateman. 1995. KPML: the KOMET-Penman (multilingual) development environment. Technical report, GMD, Darmstadt. Release 0.8.
- P. Brown and S. C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge University Press.
- J. Carletta, A. Isard, S. Isard, J. Kowtko, A. Newlands, G. Doherty-Sneddon, and A. Anderson. 1995. HCRC dialogue structure coding manual. Technical report, Association for Computational Linguistics.
- J. Ginzburg. 1995. Resolving questions, i. *Language and Philosophy*, 18(5):459–527.
- N. Green and S. Carberry. 1992. Conversational implicatures in indirect replies. In *ACL*.
- N. Green and S. Carberry. 1994. A hybrid reasoning model for indirect answers. In *ACL*.
- H. P. Grice. 1989. *Logic and Conversation*. Harvard University Press.
- M. O'Donnell. 1996. Input specification in the WAG sentence generation system. In *INLG 8*, pages 41–50.
- J. Searle. 1975. Indirect speech acts. In *Syntax and Semantics 3*. Academic Press, New York.
- A. Stenström. 1984. *Questions and Responses in English Conversation*. Lund Studies in English 68. CWK Gleerup, Malmo.
- M. Stubbs. 1983. *Discourse Analysis: The Sociolinguistic Analysis of Natural Language*. University of Chicago Press.
- P. Taylor, H. Shimodaira, S. Isard, S. King, and J. Kowtko. 1996. Using prosodic information to constrain language models for spoken dialogue. In *ICSLP 96*.