A corpus-based taxonomy of question responses

Paweł Łupkowski*
Institute of Psychology
Adam Mickiewicz University, Poznan
Pawel.Lupkowski@amu.edu.pl

Jonathan Ginzburg
UFR Études anglophones
Université Paris-Diderot (Paris 7)
yonatan.ginzburg@univ-paris-diderot.fr

Abstract

In this paper we consider the issue of answering a query with a query. Although these are common, with the exception of Clarification Requests, they have not been studied empirically. After briefly reviewing different theoretical approaches on this subject, we present a corpus study of query responses in the British National Corpus and develop a taxonomy for query responses. We sketch a formal analysis of the response categories in the framework of KoS.

1 Introduction

Responding to a query with a query is a common occurrence, representing on a rough estimate more than 20% of all responses to queries found in the British National Corpus.[1] Research on dialogue has long recognized the existence of such responses. However, with the exception of one of its subclasses—albeit a highly substantial one—the class of query responses has not been characterized empirically in previous work. The class that has been studied in some detail are Clarification Requests (CRs) (Rodriguez and Schlangen, 2004; Rieser and Moore, 2005). However, CRs can be triggered by any utterance, interrogative or otherwise. Researchers on the semantics and pragmatics of questions (see e.g. Carlson, 1983; Wiśniewski, 2003) have been aware for many years of the existence of one class of query responses—responses that express questions dependent on the question they respond to, as in (1a,b). This lead Carlson to propose (1c) as a sufficient condition for a query response, which can be formalized using (1d), assuming notions of resolvedness and aboutness (for which see e.g. Ginzburg and Sag, 2000).

(1) a. **A:** Who murdered Smith? **B:** Who was in town?

b. **A:** Who is going to win the race? **B:** Who is going to participate?

c. Who killed Smith depends on who was in town at the time.

d. q_2 can be used to respond to q_1 if q_1 depends on q_2.

e. q_1 depends on q_2 iff any proposition p such that p Resolves q_2, also satisfies p entails r such that r is About q_1.

Larsson (2002) and Asher and Lascarides (2003) argue that the proper characterization of question responses is pragmatically based. Asher and Lascarides (2003) propose to characterize non-CR query responses by means of the rhetorical relation *question elaboration* (Q-Elab) with stress on the plan-oriented relation between the initial question and the question expressed by the response. Q-Elab might be informally summarized as follows:

(2) If Q-Elab(α, β) holds between an utterance α uttered by A, where g is a goal associated by convention with utterances of the type α, and the question β uttered by B, then any answer to β must elaborate a plan to achieve g.

*The author was supported by the Iuventus Plus grant (IP2011-031-771).

[1] In the spoken part of the BNC, using SCoRE (Purver, 2001), we found 11312 ?/? cross-turn sequences, whereas 41041 ?/. cross-turn sequences, so the ?/? pairs constitute 21.6%. (For the SCoRE syntax see http://www.dcs.qmul.ac.uk/imc/ds/score/help.html.)

Table 1: Tags used to annotate question—question-response sample

Tag	Question-response category
CR	clarification requests
DP	dependent questions
FORM	questions considering the way of answering $q1$
MOTIV	questions about the underlying motivations behind asking $q1$
NO ANSW	questions aimed at avoiding answering $q1$
QA	questions providing an answer to $q1$
IGNORE	questions ignoring $q1$
IND	questions with a presupposed answer

The relation of Q-Elab, motivated by interaction in cooperative settings, is vulnerable to examples such as those in (3). (3a) has one understanding that might be characterized using dependence (*What I like depends on what* YOU *like*), but can also be used simply as a coherent retort. (3b) could possibly be used in political debate without it necessarily involving an attempt to discover an answer to the first question asked.

(3) a. **A:** What do you like? **B:** What do you like?

b. **A:** What is Sarkozy going to do about it? **B:** What is Papandreou?

In order to better understand the nature of question responses, we ran a corpus study on the British National Corpus (BNC). The results we obtained show that, apart from CRs, dependent questions are indeed by far the largest class of question responses. However, they reveal also the existence of a number of response categories, characterizable neither as dependent questions nor as plan supporting responses. They include (a) a class akin to what Conversation Analysts refer to as *counters* (Schegloff, 2007)—responses that attempt to foist on the conversation a distinct issue from the current discourse topic and (b) responses that ignore the current topic but address the situation it concerns.

Attaining completeness in characterizing the response space of a query is of fundamental importance for dialogue management and the design of user interfaces. Beyond that general goal, a better understanding of e.g. *counters* and *situation–relevant responses*, which we believe are rare in task–oriented dialogue, is important for adversarial interaction (courtroom, interrogation, argumentation, certain games). Characterizing their coherence is challenging for all approaches that ground dialogue on cooperativity principles (e.g. Asher and Lascarides, 2003; Roberts, 2011).

The rest of the paper is structured as follows: in section 2 we present the taxonomy underlying our corpus study; section 3 describes the results; in section 4 we sketch a formal analysis of one of the response categories in the framework of KoS (Ginzburg and Fernández, 2010). We conclude with a summary and future work.

2 A corpus-based taxonomy of answering by means of questions

The study sample The taxonomy of query responses was designed after an analysis of 1051 examples of query-query response pairs obtained from the BNC. The sample was obtained from blocks D, F, G, H, J, K of the BNC (so it covers a wide range of dialogue domains, like interviews, radio and TV broadcasts, tutorials, meetings, training sessions or medical consultations). Initially, examples were obtained with the search engine SCoRE (Purver, 2001) (the search string was `?$ | ?$`). Subsequently, cross talk and tag questions were eliminated manually. The sample was classified and annotated by the first author with tags presented in Table 1 (we discuss the reliability of this annotation in section 3).

In what follows we describe and exemplify each class of the resulting taxonomy. To make the description clear we will use $q1$ for the initial question posed and $q2$ for a question given as a response to $q1$. The taxonomy was built with attention paid to the function of $q2$ in the dialogue.

Clarification requests (CR) Clarification requests are all question-responses that concern the content or form of $q1$ that was not completely understood. This class contains intended content queries (4a),

repetition requests (4b) and relevance clarifications (4c).

(4) a. **A:** What's Hamlet about? **B:** Hamlet? [KPW, 945–946][2]

b. **A:** Why are you in? **B:** What? **A:** Why are you in? [KPT, 469–471]

c. **A:** Is he knocked out? **B:** What do you mean? [KDN, 3170–3171]

In this paper we will not consider this class in detail, mainly because of existing, detailed work on this subject such as (Purver, 2006).

Dependent questions (DP) By a *dependent question* we understand $q2$ where a dependency statement as in (1c) could be assumed to be true. The following examples illustrate this:

(5) a. **A:** Do you want me to <*pause*> push it round? **B:** Is it really disturbing you? [FM1, 679–680]
(cf. *Whether I want you to push it depends on whether it really disturbs you*

b. **A:** Any other questions? **B:** Are you accepting questions on the statement of faith at this point? [F85, 70–71]
(cf. *Whether more questions exist depends on whether you are accepting questions on the statement of faith at this point.*)

'How should I answer this?' questions (FORM) This class consists of question-responses addressing the issue of the *way* the answer to $q1$ should be given. It is the case where the *form* of answer to $q1$ depends on the answer given to $q2$. This relation between $q1$ and $q2$ might be noticed in following examples. Consider (6a). The way B answers A's question in this case will be dictated by A's answer to $q2$—whether or not, A wants to know details point by point.

(6) a. **A:** Okay then, Hannah, what, what happened in your group?
B: Right, do you want me to go through every point? [K75, 220–221]

b. **A:** Where's that one then?
B: Erm, you know Albert Square? [KBC, 506–507]

Requests for underlying motivation (MOTIV) In the case of *requests for underlying motivation* $q2$ addresses the issue of motivation behind asking $q1$. What is important here is that the fact of answering $q1$ depends on the answer to $q2$ (i.e. providing proper reasons for asking $q1$). In this aspect this class differs form the previous ones, where we may assume that an agent wishes to provide answer to $q1$. Most of question-responses of this kind are of the form "Why?" (32 out of 41 gathered examples, cf. (7a)), but also other formulations were observed (8 out of 41, cf. (7b)). Most direct questions of this kind are: *What's it got to do with you?*; *what's it to you?*; *Is that any of your business?*; *what's that gotta do with anything?*.

(7) a. **A:** What's the matter? **B:** Why? [HDM, 470–471]

b. **A:** Out, how much money have you got in the building society? **B:** What's it got to do with you? [KBM, 2086–2087]

I don't want to answer your question (NO ANSW) The role of query responses of this class is to give a clear signal that an agent does not want to provide an answer to $q1$. Instead of answering $q1$ the agent provides $q2$ and attempts to 'turn the table' on the original querier.

(8) a. **A:** Yeah what was your answer? **B:** What was yours? [KP3, 636–637]

b. **A:** Why is it recording me? **B:** Well why not? [KSS, 43–44]

[2]This notation indicates the BNC file (KPW) together with the sentence numbers (945–946).

Table 2: Frequency of question—question-response categories. The parenthesized percentage is the category's percentage of the sample that *excludes* CRs.

Category	Frequency	% of the Total
CR	832	79.16
DP	108	10.28 (49)
MOTIV	41	3.90 (18)
NO ANSW	26	2.47 (12)
FORM	16	1.52 (7)
QA	13	1.24 (6)
IND	9	0.85 (4)
IGNORE	6	0.57 (3)
Total	**1051** (219)	100

Indirect answers/responses (IND/QA) This class consists of query responses, which provide (indirectly) an answer to $q1$. Interestingly, answering $q2$ is not necessary in this case. Consider (9a). B by asking the question *Do you know how old this sweater is?* clearly suggests that the answer to A's question is negative. Moreover, B does not expect to obtain an answer to his/her question. This might also be observed in (9b) ('of course I am Gemini').

(9) a. **A:** Is that an early Christmas present, that sweater? **B:** Do you know how old this sweater is? [HM4, 7–8]

b. **A:** Are you Gemini? **B:** Well if I'm two days away from your, what do you think? [KPA, 3603–3604]

Another means of providing indirect answers can be observed in the corpus data. It is the case that by asking $q2$ an agent already presupposes the answer to $q1$. If we take a look on (10) we note that positive answer to $q1$ is presupposed in B's question (I will help you).

(10) **A:** Will you help with the <*pause*> the paint tonight? **B:** What can I do? [KE4, 3263–3264]

I ignore your question (IGNORE) The last observed class is somewhat harder to grasp. This is the case where $q2$ is related to the situation, but ignores $q1$. This is evident in (11). A and B are playing *Monopoly*. A asks a question, which is ignored by B. It is not that B does not want to answer A's question and that's why he/she asks $q2$. Rather, B ignores $q1$ and asks a question related to the situation (in this case the board game).

(11) **A:** I've got Mayfair <*pause*> Piccadilly, Fleet Street and Regent Street, but I never got a set did I?
B: Mum, how much, how much do you want for Fleet Street? [KCH, 1503–1504]

3 Results and annotation reliability

The results of the performed classification are presented in Table 2. Putting aside CRs, the majoritarian class is indeed DP. What is striking is the relatively large frequency of adversarial responses (the classes MOTIV, NO ANSW, IGNORE). FORM, as we discuss below, is the sole class whose coherence clearly requires reasoning about the querier's intentions. It is relatively infrequent.

In order to check the reliability of the classification process, the decision tree was tested by three other annotators. Annotators obtained the sample of 90 (randomly chosen) question-question pairs[3] and decision tree. The instruction was to annotate question-reply to the first question in each example. Some of the examples were enriched with additional context (after q2). Two annotators reported that the annotation task would be easier if the context would be present for all examples.

The reliability of the annotation was evaluated using κ (Carletta, 1996). The agreement on the coding of the control sample by four annotators was moderate (Fleiss $\kappa = 0.44$, $SE = 0.0206$, $95\%CI =$

[3]The distribution of the classes was in line with the distribution observed by the primary annotator: CR: 39 examples; DP: 18 examples; MOTIV: 10 examples; NO ANSW: 10 examples; FORM: 4 examples; QA: 4 examples; IGNORE: 3 examples; OTHER: 2 examples.

0.3963 to 0.4770)[4]. One of the control sample annotators is an experienced linguist with extensive past work with dialogue transcripts. In this case agreement on the coding was strong (71% agreement with Cohen's $\kappa = 0.62$, $SE = 0.0637$, $95\%CI = 0.4902$ to 0.7398). Two other control sample annotators are logicians, but with little experience in corpus annotation. For them agreement on the coding was somewhat lower, i.e. moderate (66% agreement with Cohen's $\kappa = 0.56$, $SE = 0.0649$, $95\%CI = 0.4266$ to 0.6810; and 54% agreement with Cohen's $\kappa = 0.42$, $SE = 0.0674$, $95\%CI = 0.2829$ to 0.5472). The most unproblematic cases were `CR, MOTIV` and `IGNORE` (the largest groups of examples with at least 3 annotators' agreement). Also `DP, NO ANSW` and `QA` had high agreement between annotators. The main problem was with `FORM`. We assume that this is caused by the unclarity in the question introducing this class in the decision tree ('The way the answer to $q1$ will be given depends on the answer to $q2$', while for DP it was 'Is it the case that the answer to $q1$ depends on the answer to $q2$?'). Feedback from two of three control sample annotators reported this as a confusing case. There were two cases in the control sample on which annotators completely disagreed. These were the following:

(12) a. **A:** You know the one you just took out? **B:** You want it back? [F77, 86–87]

b. **A:** You want a drink dear? **B:** Have your sweets for what? [KD1, 5132–5133]

4 Modeling Query Response Categories in KoS

In this section we show how to explicate the coherence relation that underlies the `DP` query responses within the framework of KoS. It is worth mentioning that this framework allows to model also the other query responses types described in this article, as we will show in an extended version of this paper. KoS is a framework for dialogue whose logical underpinning is Type Theory with Records (TTR) (Cooper, 2005) and which underlies dialogue systems such as GoDiS and CLARIE (Larsson, 2002; Purver, 2006). On the approach developed in KoS, there is actually no single context—instead of a single context, analysis is formulated at a level of information states, one per conversational participant. The type of such information states is given in (13a). We leave the structure of the private part unanalyzed here, as with one exception all our characterizations do not make reference to this; for one approach to $private$, see e.g. (Larsson, 2002). The dialogue gameboard represents information that arises from publicized interactions. Its structure is given in (13b)—the *spkr,addr* fields allow one to track turn ownership, *Facts* represents conversationally shared assumptions, *Pending* and *Moves* represent respectively moves that are in the process of/have been grounded, *QUD* tracks the questions currently under discussion:

(13) a. TotalInformationState (TIS) $=_{def}$

$$\begin{bmatrix} \text{dialoguegameboard} : \text{DGB} \\ \text{private} : \text{Private} \end{bmatrix}$$

b. DGBType $=_{def}$

$$\begin{bmatrix} \text{spkr: Ind} \\ \text{addr: Ind} \\ \text{utt-time} : \text{Time} \\ \text{c-utt} : \text{addressing(spkr,addr,utt-time)} \\ \text{Facts} : \text{Set(Proposition)} \\ \text{Pending} : \text{list(locutionary Proposition)} \\ \text{Moves} : \text{list(locutionary Proposition)} \\ \text{QUD} : \text{poset(Question)} \end{bmatrix}$$

The basic units of change are mappings between dialogue gameboards that specify how one gameboard configuration can be modified into another on the basis of dialogue moves. We call a mapping between DGB types a *conversational rule*. The types specifying its domain and its range we dub, respectively, the *preconditions* and the *effects*, both of which are supertypes of DGBType.

We start by characterizing the moves that typically involve accepting $q1$ into the DGB. The potential for `DP` responses is explicated on the basis of the two conversational rules in (14a,b): (14a) says that given a question q and ASK(A,B,q) being the LatestMove, one can update QUD with q as QUD–maximal.

[4] All the data established with `http://www.stattools.net`. Access 25.11.2012.

QSPEC is what characterizes the contextual background of reactive queries and assertions. (14b) says that if q is QUD–maximal, then subsequent to this either conversational participant may make a move constrained to be q–specific (14c):

(14) a. Ask QUD–incrementation

$$\left[\begin{array}{l}\text{pre}: \left[\begin{array}{l}\text{q}: \text{Question} \\ \text{LatestMove} = \text{Ask(spkr,addr,q)}: \text{IllocProp}\end{array}\right] \\ \text{effects}: \left[\text{qud} = \left\langle \text{q,pre.qud} \right\rangle : \text{poset(Question)}\right]\end{array}\right]$$

b. QSPEC

$$\left[\begin{array}{l}\text{pre}: \left[\text{qud} = \left\langle \text{q, Q} \right\rangle : \text{poset(Question)}\right] \\ \text{effects}: \text{TurnUnderspec} \wedge_{merge} \\ \left[\begin{array}{l}\text{r}: \text{AbSemObj} \\ \text{R}: \text{IllocRel} \\ \text{LatestMove} = \text{R(spkr,addr,r)}: \text{IllocProp} \\ \text{c1}: \text{Qspecific(r,q)}\end{array}\right]\end{array}\right]$$

c. q-specific utterance: an utterance whose content is either a proposition p About q or a question q_1 on which q Depends

5 Summary and Future Work

The paper provides the first empirically-based study of query responses to queries. The most interesting finding here is the existence of a number of classes of adversarial responses, that involve the rejection/ignoring of the original query. Indeed, in such cases the original query is rarely responded to in subsequent interaction.

We conducted our study in the BNC since it is a general corpus with a variety of domains and genres. It is of course important to extend this study to more detailed consideration of specific genres and domains. A significant challenge for future work is automatic classification of query responses into a taxonomy like the one provided here. We intend to address this in future work.

References

Asher, N. and A. Lascarides (2003). *Logics of Conversation*. Cambridge: Cambridge University Press.

Carletta, J. (1996). Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics 22*(2), 249–254.

Carlson, L. (1983). *Dialogue Games*. Synthese Language Library. Dordrecht: D. Reidel.

Cooper, R. (2005). Austinian truth in Martin-Löf type theory. *Research on Language and Computation 3*(4), 333–362.

Ginzburg, J. and R. Fernández (2010). Computational models of dialogue. In A. Clark, C. Fox, and S. Lappin (Eds.), *Handbook of Computational Linguistics and Natural Language*, Oxford. Blackwell.

Ginzburg, J. and I. A. Sag (2000). *Interrogative Investigations: the form, meaning and use of English Interrogatives*. Number 123 in CSLI Lecture Notes. Stanford: California: CSLI Publications.

Larsson, S. (2002). *Issue based Dialogue Management*. Ph. D. thesis, Gothenburg University.

Purver, M. (2001, October). SCoRE: A tool for searching the BNC. Technical Report TR-01-07, Department of Computer Science, King's College London.

Purver, M. (2006). Clarie: Handling clarification requests in a dialogue system. *Research on Language & Computation 4*(2), 259–288.

Rieser, V. and J. Moore (2005). Implications for generating clarification requests in task-oriented dialogues. In *Proceedings of the 43rd Meeting of the Association for Computational Linguistics*, Michigan.

Roberts, C. (2011). Information structure: Afterword. *Semantics and Pragmatics*.

Rodriguez, K. and D. Schlangen (2004). Form, intonation and function of clarification requests in german task-oriented spoken dialogues. In J. Ginzburg and E. Vallduvi (Eds.), *Proceedings of Catalog'04, The 8th Workshop on the Semantics and Pragmatics of Dialogue*, Universitat Pompeu Fabra, Barcelona.

Schegloff, E. (2007). *Sequence Organization in Interaction*. Cambridge: Cambridge University Press.

Wiśniewski, A. (2003). Erotetic search scenarios. *Synthese 134*, 389–427.