Query responses

Paweł Łupkowski[1] *and Jonathan Ginzburg*[2]
[1] Department of Logic and Cognitive Science
Institute of Psychology, Adam Mickiewicz University
Poznań, Poland
[2] Laboratoire de Linguistique Formelle (LLF) (UMR 7110) & LabEx-EFL
Université Paris-Diderot (Paris 7)
Sorbonne Paris-Cité

ABSTRACT

In this article we consider the phenomenon of answering a query with a query. Although such answers are common, no large-scale, corpus-based characterization exists, with the exception of clarification requests. 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 identify a variety of response categories that have not been formalized in previous dialogue work, particularly those relevant to adversarial interaction. We show that different response categories have significantly different rates of subsequent answer provision. We provide a formal analysis of the response categories within the framework of KoS.

Keywords: questions, query responses, corpus study, 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 (BNC).[1] Research on dialogue has long recognized the existence of such responses. However, with

[1] In the spoken part of the BNC, using SCoRE (Purver 2001), we found 9,279 ?/? cross-turn sequences, whereas 41,041 ?/. cross-turn sequences, so the ?/? pairs constitute 22.61%.

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 is that of Clarification Requests (hereafter referred to as CRs) (see e.g., Purver *et al.* 2001; Rodriguez and Schlangen 2004; Rieser and Moore 2005). However, CRs can be triggered by any utterance, interrogative or otherwise. Researchers working on the semantics and pragmatics of questions (see e.g., Carlson 1983; Wiśniewski 1995) have been aware for many years of the existence of one class of query responses – responses that express questions dependent in some sense on the question they respond to, as in (1a,b). This led Carlson to propose (1d) as a sufficient condition for a query response (cf. (1a,c)).

(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. $q2$ can be used to respond to $q1$ if $q1$ depends on $q2$.

How to define question dependence is an important issue if the criterion in (1d) is to have much substance. A number of proposals concerning dependence have been made in the literature, for instance Ginzburg (2012) offers the definition in (2):

(2) $q1$ depends on $q2$ iff any proposition p such that p resolves $q2$, also satisfies p entails r such that r is about $q1$. (Ginzburg 2012, (61b), p. 57)

For Ginzburg, this notion of dependence is an agent-relative notion, given the agent-relativity of the relation resolves.[2] An arguably more open-ended view is taken by Roberts (1996), who suggests that a query move m is relevant in a context where q is the question under discussion if *m is part of a strategy to answer q* (Roberts 1996, p. 17). In similar fashion, Larsson (2002) and Asher and Lascarides (2003) argue

[2] The agent-relativity of the relation resolves is argued for in great detail in Ginzburg 1995. resolves is the answerhood notion implicated in examples such as '… knows where she is' and '… knows who came to the talk', which is, arguably, relativized by agent goals and background knowledge.

that the proper characterization of query responses is pragmatically based. Asher and Lascarides (2003) propose to characterize non-CR query responses by means of the rhetorical relation of *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:

(3) 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.

Q-Elab, motivated by interaction in cooperative settings, is vulnerable to examples such as those in (4). There is a reading of (4a) that can be characterized by using dependence (*What I like depends on what* YOU *like*), but it can also be used simply as a coherent retort. (4b) could possibly be used in political debate without necessarily involving any attempt to discover an answer to the first question

(4) a. **A:** What do you like? **B:** What do you like?
b. **A:** What is Sarkozy going to do about it? **B:** Well, what is Papandreou going to do?

In the field of the logic of questions we can mention approaches proposed within Inferential Erotetic Logic (IEL) (Wiśniewski 1995, 2013) and inquisitive semantics (INQ) (Groenendijk 2009; Groenendijk and Roelofsen 2011). Although INQ and IEL represent different approaches to questions, both frameworks share a similar treatment of question dependency. In IEL, the central notion used to express dependency between questions is *erotetic implication*. Erotetic implication is a semantic relation between a question, Q, a (possibly empty) set of declarative well-formed formulae, X, and a question, Q_1. Intuitively, erotetic implication ensures the following: (i) if Q has a true direct answer and X consists of truths, then Q_1 has a true direct answer as well ('transmission of soundness[3] and truth into soundness' – cf. Wiśniewski 2003, p. 401), and (ii) each direct answer to Q_1, if true, and if all elements of X are true, narrows down the class for

[3] A question Q is *sound* iff it has a true direct answer (with respect to the underlying semantics).

which a true direct answer to Q can be found ('open-minded cognitive usefulness' – cf. Wiśniewski 2003, p. 402).

In the framework of inquisitive semantics, the dependency relation has been analysed in terms of *compliance*. Roughly speaking, INQ treats questions as sets of possibilities or, in other words, as an issue to be resolved. The intuition behind the notion of compliance is to provide a criterion to "judge whether a certain conversational move makes a significant contribution to resolving a given issue" (Groenendijk and Roelofsen 2011, p. 167).

Other question generation mechanisms in a broadly dialogical context have been proposed in the literature. One such notion is *askability*. The intuition behind askability relates to the issue – when is it reasonable to (publicly) ask a question? Peliš and Majer (2010), applying a dynamic epistemic logic of questions combined with a public announcements' logic for modelling communicative interaction and knowledge revision during this process, propose three conditions that have to be met by an agent in order to ask a question within a group of agents: (i) the answer is not known to the agent posing the question (non-triviality); (ii) each direct answer is considered as possible by the agent (admissibility); and (iii) at least one of the direct answers must be the right one in a given context (context condition).

Van Kuppevelt (1995) proposes *topicality* as the general organizing principle in discourse. The topic (for a discourse unit) is provided by an explicit or implicit question. Van Kuppevelt does not consider simple question – query response pairs, but rather speaks about discourse units. However, the relation between such units is determined by the relation between the previously mentioned topic-providing questions. From the current perspective, the most interesting is the notion of *subtopic-constituting sub-question*:

(5) An explicit or implicit question Q_p is a subtopic-constituting subquestion if it is asked as the result of an unsatisfactory answer A_{p-n} to a preceding question Q_{p-n} and is intended to complete A_{p-n} to a satisfactory answer to Q_{p-n}. (Kuppevelt 1995, p. 125)

Graesser *et al.* (1992) propose four question generation mechanisms for natural settings (especially in educational contexts). The first group consists of *knowledge deficit* questions. The other three groups are: *common ground* questions, *social coordination* questions and *conversation control* questions. Common ground questions, like 'Are we working on the third problem?' or 'Did you mean the independent variable?', are asked to check whether knowledge is shared between dialogue participants. Social coordination questions relate to different roles of dialogue participants, such as in student – teacher conversations. Social coordination questions are requests for permission to perform a certain action or might be treated as indirect request for the addressee to perform such an action (e.g., 'Could you graph these numbers?', 'Can we take a break now?'). Conversation control questions, as it is indicated by their name, aim at manipulating the flow of a dialogue or the attention of its participants (e.g., 'Can I ask you a question?').

How many kinds of query responses are there and what aspects of context or agents' information states are needed to characterise them? In order to better understand the nature of query response, we ran a corpus study on one large, balanced corpus, the British National Corpus (BNC), and several smaller, more domain-specific corpora, a selection from CHILDES (parent/child interaction; MacWhinney 2000), AMEX (interactions in the travel domain; Kowtko and Price 1989), and BEE (tutor/student interaction; Rosé *et al.* 1999). The results we obtained, discussed in Section 3 of this paper, show that, apart from CRs, dependent questions are indeed by far the largest class of query responses. However, our results reveal also the existence of a number of response categories characterisable neither as dependent questions nor as plan-supporting responses. These include:

- a class akin to what Conversation Analysts refer to as *counters* (Schegloff 2007) – responses that attempt to foist on the conversation an issue that differs from the current discourse topic and
- *situation-relevant responses* – responses that ignore the current topic but address the situation it concerns.

Just as wide coverage is an important goal for any computational theory of sentential grammar (tempered by some notion of 'strong generative capacity', i.e., attaining this in a principled way), the same goal *mutatis mutandis* applies to theories of dialogue; their corresponding aim is to characterise in a principled way the relevance or coherence of a wide range of utterance sequences. Attaining wide coverage for the particular case of the response space of a query naturally has significant practical 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* is important for adversarial interaction (e.g., courtroom, interrogation, argumentation, certain games).

The rest of the paper is structured as follows: in Section 2 we present the taxonomy underlying our corpus study; Sections 3 and 4 describe the results, whereas issues concerning annotation reliability are discussed in Section 5; in Section 6 we show how to analyse the relevance of each of the response categories emergent in the corpus study. We do this in terms of information state transitions of two interlocutors participating in a dialogue, using the dialogue framework, KoS.[4] We conclude with a brief cross-theoretical evaluation of potential analyses of the various response classes, and with possibilities for future work.

2 A CORPUS-BASED TAXONOMY OF QUERY RESPONSES

In this section, we present a corpus-based taxonomy of query responses. It was designed on the basis of 1,051 examples of query – query-response pairs obtained from the BNC. Initially, examples were obtained using the search engine SCoRE (Purver 2001). Subsequently, cross talk and tag questions were eliminated manually. The annotation was performed by the first author; we discuss the reliability of this annotation in Section 5. In what follows, we describe and exemplify each class of the resulting taxonomy. To make the description clearer we use $q1$ for the initial question and $q2$ for a question given as a response to $q1$. The taxonomy is focused on the *function of* $q2$ in a dialogue.

2.1 *Clarification requests (CR)*

Clarification requests are all query responses that concern any aspect pertaining to the content or form of $q1$ that was not completely un-

[4] KoS is a toponym – the name of an island in the Dodecanese archipelago – bearing a loose connection to *conversation-oriented semantics* (Ginzburg 2012, p. 2).

derstood. This class contains intended content queries (see example (6a)), repetition requests (example (6b)) and relevance clarifications (example (6c)). In this article, we will not consider this class in detail, mainly because of existing, detailed work on this subject (see e.g., Purver 2006; Ginzburg 2012).

(6) a. **A:** What's Hamlet about?
B: Hamlet? [KPW, 945–946][5]

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]

2.2 *Dependent questions (DP)*

By a *dependent question,* we understand $q2$ where a dependency statement as in (1d) (see page 246) could be assumed to be true. The following examples illustrate this:

(7) 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 around 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.*)

c. **A:** Does anybody want to buy an Amstrad? <*pause*>
B: Are you giving it away? [KB0, 3343–3344]
(cf. *Whether anybody wants to buy an Amstrad depends on whether you are giving it away.*)

2.3 *'How should I answer this?' questions (FORM)*

This class consists of query responses addressing the issue of the *way* the answer to $q1$ should be given. In other words, whether the answer

[5] This notation indicates the sentence numbers (945–946) of a BNC file (KPW).

to $q1$ will be satisfactory to A depends on $q2$. This relation between $q1$ and $q2$ is illustrated in the following examples. Consider (8a). 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.

(8) 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]

c. **A:** Another thing I found out today was do we know where our main supplier of our coffee is.
Any guesses?
B: Which country? [G3U, 251–253]

2.4 *Requests for underlying motivation (MOTIV)*

In the case of *requests for underlying motivation*, $q2$ addresses the issue of the motivation underlying asking $q1$. Whether an answer to $q1$ will be provided depends very much on receiving a convincing answer to $q2$ (i.e., one that provides good reasons for asking $q1$). In this respect this class differs from the previous classes, where we may assume that an agent wishes to provide an answer to $q1$. Most query responses of this kind are of the form 'Why?' (32 out of 41 examples, see e.g., (9a)) but also other formulations were observed (9 out of 41, see e.g., (9b) and (9c)). 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?*.

(9) 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]

c. **A:** Just what the fucking do you think you're doing?
B: Is that any of your business? [KDA, 1308–1309]

2.5 *'I don't want to answer your question' (NO ANSW)*

The role of query responses of this class is to signal that an agent does not want to provide an answer to $q1$, at least at the current stage of the conversation. Instead of answering $q1$, the agent provides $q2$ and attempts to "turn the table" on the original querier, as in examples (10).

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

b. **A**: come on Stacey get on with it <*pause*> can you move up a bit?
B: What? <*unclear*> why didn't you pull the bench out? [KCG, 378–379]

c. **A**: What about my fifty p?
B: Fucking hell, where's my tenner? [KDA, 3527–3528]

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

2.6 *Indirect responses (IND)*

This class consists of query responses, which provide (indirectly) an answer to $q1$. Interestingly, providing an answer to $q2$ is not necessary in this case. Consider (11a). By asking the question *Do you know how old this sweater is?*, B 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 may also be observed in examples (11b) (*of course I am Gemini*) and (11c) (*no, my job is not safe*).

(11) 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]

c. **A:** Is your job safe?
B: Well, whose job's safe? [G5L, 130–131]

Another means of providing indirect answers can also be observed in the corpus data. These are cases where by asking $q2$ an agent already presupposes the answer to $q1$. (12a) illustrates this – we note that a positive answer to $q1$ is presupposed in B's question. A similar situation can be observed in examples (12b) (*no, I have not tasted this*) and (12c) (*I will help you*).

(12) a. **A:** I've got to do the washing up?
B: Shall I, shall I come and help you? [KPU, 1861–1862]

b. **A:** have you tasted this?
B: are they nice? [KPY, 653–654]

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

2.7 *'I ignore your question' (IGNORE)*

The final class in the taxonomy involves cases where $q2$ does not address $q1$, but is, nonetheless, related to the situation associated with $q1$. This is evident in example (13c). A and B are playing Monopoly. A asks a question, which is ignored by B. It is not that B does not wish to answer A's question and therefore asks $q2$. Rather, B ignores $q1$ and asks a question related to the situation (in this case, the board game).

(13) a. **A:** Well do you wanna go down and have a look at that now? <*pause*> While there's workmen there?
B: Why haven't they finished yet? [KCF, 617–619]

b. **A:** Just one car is it there?
B: Why is there no parking there? <*unclear*> [KP1, 7882–7883]

c. **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]

2.8 *Summary*

In this section, a corpus-based taxonomy of query responses was presented. Seven classes of query-responses were described. The classification is focused on the function $q2$ (the question given as a response) serves in relation to $q1$ (the initial question). In what follows, we present the corpus study that led to the classification. First, a study using the BNC is discussed, then the class distribution over specific genres is presented. Subsequently, we consider the issue of annotation reliability.

3 RESULTS

As we noted, this study used a sample of 1,051 query – query response pairs from the BNC. The procedure for obtaining the sample was the following. First the search engine SCoRE was used on the whole spoken part of the BNC using as the search string: `? $ | ? $`.[6] Following this, the search results were checked manually. The collected sample covers a wide range of dialogue domains, including interviews, radio and TV broadcasts, tutorials, meetings, training sessions or medical consultations (blocks D, F, G, H, J, and K of the BNC). The summary of dialogue domains for the sample is presented in Table 1.

Domain	Frequency	% of the Total
free conversation	940	89.44
educational context (lesson, tutorial, training)	36	3.43
meeting (public meeting, seminar, conference)	27	2.57
radio broadcast	25	2.38
interview	15	1.43
medical consultation	4	0.38
TV broadcast	4	0.38
Total	1,051	100

Table 1: Dialogue domains in the research sample (BNC)

The sample was classified and annotated by the first author with the tags presented in Table 2.

Tag	Query-response type
`CR`	clarification requests
`DP`	dependent questions
`FORM`	questions considering means of answering $q1$
`MOTIV`	questions about the motivations underlying asking $q1$
`NO ANSW`	questions aimed at avoiding to answer $q1$
`IND`	questions with a presupposed answer
`IGNORE`	questions ignoring $q1$

Table 2: Tags used to annotate the query – query response sample

[6] The expression '? $' matches any sentence/turn with a question mark at the end and the pipe character matches the break between sentences/turns. For more details about the SCoRE syntax see `http://www.dcs.qmul.ac.uk/imc/ds/score/help.html`.

Table 3:
Frequency of query – query response categories in the BNC (The parenthesized percentage is the percentage recalculated once the CRs are excluded from the sample.)

	Category	Frequency	% of the Total	
1.	CR	832	79.16	
2.	DP	108	10.28	(49.31)
3.	MOTIV	41	3.90	(18.72)
4.	NO ANSW	26	2.47	(11.87)
5.	FORM	16	1.52	(7.31)
6.	IND	22	2.09	(10.05)
7.	IGNORE	6	0.57	(2.74)
	Total	**1,051** (219)	100	

To guide the classification process, we used the following questions:

1. (CR) Is $q2$ a query about something not completely understood in $q1$?
2. (DP) Is it the case that the answer to $q1$ depends on the answer to $q2$?
3. (MOTIV) Does $q2$ address the motivation underlying asking $q1$?
4. (NO ANSW) Is it the case that $q2$ enables the speaker to avoid answering $q1$ while attempting to force the other speaker to answer $q2$ first?
5. (FORM) Is it the case that the way the answer to $q1$ will be given depends on the answer to $q2$?
6. (IND) Is it the case that $q2$ is rhetorical and in this sense does not need to be answered and provides (indirectly) an answer to $q1$?
7. (IGNORE) Does $q2$ relate to the situation described by $q1$?

The results of the classification are presented in Table 3. The parenthesized percentage is the percentage recalculated once the CRs are excluded from the sample.

The largest class after CRs is DP. What is rather striking is the relatively large frequency of adversarial responses (the classes MOTIV, NO ANSW, and IGNORE).

We also compared which query categories lead to a subsequent answer, either about $q2$ or about $q1$. Bearing in mind that our taxonomy is focused on the function of $q2$ in a dialogue, we would expect the following results.

Category	Ans. to $q2$	Ans. to $q1$
DP	76.85	62.96
MOTIV	78.05	51.22
NO ANSW	80.77	11.54
FORM	68.75	81.25
IND	53.85	100
IGNORE	50	16.67

Table 4:
Answers provided to query responses in % of the total per category

DP Answering $q2$ should lead to answer concerning $q1$. The figures for $q1$ and $q2$ should be similar.

FORM Whether the answer to $q1$ will be useful for A depends on $q2$. $q2$ addresses only the form of the answer to $q1$, so is somewhat less important than with DP. Hence, the number of answers to $q1$ could be higher than for $q2$.

MOTIV Whether an answer to $q1$ will be provided depends on a satisfactory answer to $q2$. The numbers for $q1$ and $q2$ should be comparable, though $q1$ may be somewhat smaller.

NO ANSW Instead of answering $q1$, the agent provides $q2$ and attempts to "turn the table" on the original querier. The original querier is pressured to answer $q2$ and put $q1$ aside. Hence, the numbers for $q1$ should be significantly smaller than for $q2$.

IND $q2$ (indirectly) provides an answer to $q1$, so the latter is answered by definition. Providing an answer to $q2$ is not necessary in this case, so the numbers should be low here.

IGNORE The person posing $q2$ shows a lack of interest in $q1$, but since $q2$ relates to the situation associated with $q1$, there is some expectation that $q2$ be responded to. Thus, the numbers for $q1$ should be significantly smaller than for $q2$. Moreover, the numbers for $q2$ should also be rather low (asking $q2$ is not very cooperative).

The results of the data analysis are presented in Table 4. They are in line with the intuitions underlying the taxonomy.

4 CLASS DISTRIBUTION OVER SPECIFIC GENRES

We conducted our study using the BNC since it is a general corpus with a variety of domains and genres. However, we also wanted to

Table 5:
Frequency of query – query response categories (CHILDES) (The parenthesized percentage is the percentage recalculated once the CRs are excluded from the sample.)

Category	Frequency	% of the Total	
CR	319	88.12	
DP	11	3.04	(25.58)
MOTIV	2	0.55	(4.65)
NO ANSW	5	1.38	(11.63)
FORM	3	0.83	(6.98)
IND	5	1.38	(11.63)
IGNORE	17	4.70	(39.53)
Total	**362** (43)	100	

Table 6:
Frequency of query – query response categories (BEE) (The parenthesized percentage is the percentage recalculated once the CRs are excluded from the sample.)

Category	Frequency	% of the Total	
CR	10	22.22	
DP	28	62.22	(80)
NO ANSW	6	13.33	(17.14)
IGNORE	1	2.22	(2.86)
Total	**45** (35)	100	

check how the classes are distributed in more genre-specific corpora. To do this, we decided to study the following corpora:

- the Child Language Data Exchange System (CHILDES; MacWhinney 2000), which contains adult-child conversations,
- the Basic Electricity and Electronics Corpus (BEE; Rosé *et al.* 1999), which contains tutorial dialogues from electronics courses,
- the SRI/CMU American Express dialogues (AMEX; Kowtko and Price 1989), which contains conversations with travel agents.

As with the BNC study, the data was initially obtained by using the search engine SCoRE. Subsequently, cross talk and tag questions were eliminated manually. The annotation was then performed by the first author. 362 examples were obtained from the sample of the CHILDES corpus (files *bates*, *belfast*, and *manchester/anne*); 45 examples were obtained from the whole BEE corpus and 8 from the whole AMEX corpus (the low numbers for BEE and AMEX are caused by the significantly smaller size of these corpora in comparison to BNC and CHILDES). The results of the classification applied to these corpora are presented in Tables 5, 6 and 7. The parenthesized percentage is the percentage recalculated once the CRs are excluded from the sample.

Category	Frequency	% of the Total
CR	1	12.5
DP	7	87.5 (100)
Total	**8** (7)	100

Table 7:
Frequency of query – query response categories (AMEX) (The parenthesized percentage is the percentage recalculated once the CRs are excluded from the sample.)

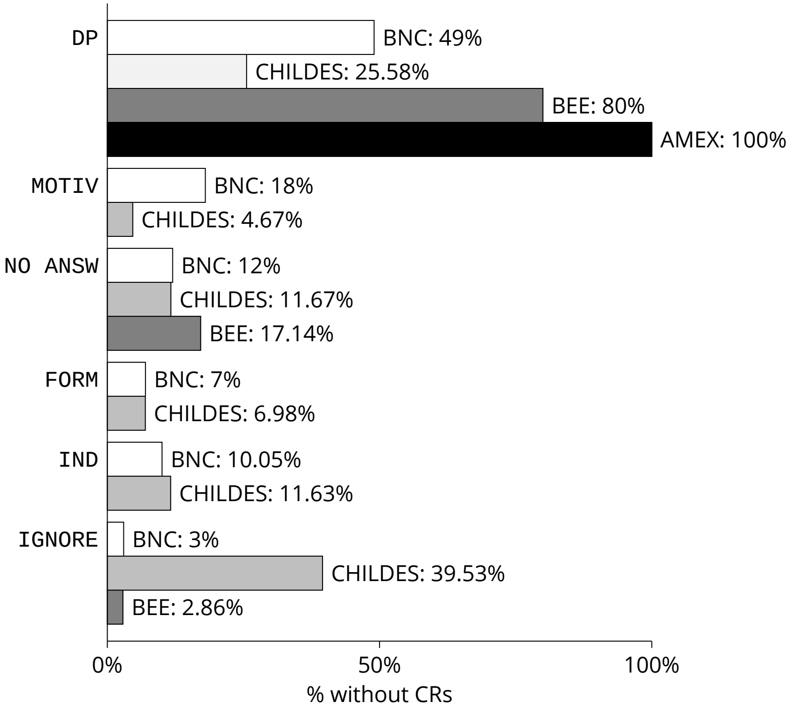
As is readily apparent, the DP class is the second largest class in the CHILDES corpus and is the largest class in the task-oriented dialogues obtained from the BEE and AMEX corpora. As for the adversarial classes (MOTIV, NO ANSW, IGNORE), these are very rare in task oriented dialogues. One exception is the NO ANSW class in the case of the BEE corpus. Here the percentage of NO ANSW questions is even higher than in the BNC and in CHILDES. This type of query response is used in a teaching context to encourage a student to provide his/her answer to the teacher's question (e.g., Student: can you remind me which colors mean what on the different resistors?; Tutor: Is that the first thing you need to know? [log-stud31]). When it comes to the CHILDES corpus, a large percentage of IGNORE query responses was observed – in all the examples, it was a child who used this kind of query response. One can also note that for NO ANSW, FORM and IND, the frequency is similar for CHILDES and the BNC. The summary of the distributions for the BNC, CHILDES, AMEX and BEE is presented in Figure 1.

We also compared which query categories lead to a subsequent answer, either about $q2$ or about $q1$. The results are presented in Table 8. In terms of answer analysis, task-oriented dialogues are inter-

Table 8: Answers provided to query responses (CHILDES, BEE and AMEX) in % of the total

	CHILDES		BEE		AMEX	
Category	Ans. to $q2$	Ans. to $q1$	Ans. to $q2$	Ans. to $q1$	Ans. to $q2$	Ans. to $q1$
DP	72.73	45.45	96.43	96.43	100	100
MOTIV	50	0	——	——	——	——
NO ANSW	80	20	100	50	——	——
FORM	100	100	——	——	——	——
IND	0	100	——	——	——	——
IGNORE	70.59	5.88	0	0	——	——

Figure 1: Summary of the distributions in each corpus (in % of the total, without CRs)



esting in the context of the `DP` query response class. For all observed examples, an answer was provided to $q2$ and to $q1$. The `NO ANSW` category also behaves in line with the observations for the BNC. We can observe the fulfilment of some of our predictions in the case of the CHILDES corpus. When it comes to interaction with children, neither maintaining attention nor topic continuity are a given, and this can be observed in the data. In the case of `DP` questions, we still have a high number of answers provided for $q2$, but the number of answers provided to $q1$ is relatively low. As for the `IGNORE` class, our prediction in general was that the number of answers provided for $q2$ should be low (since the behavior it represents is not very cooperative). However, for CHILDES we observe a high number of provided answers. In our sample it was a child who posed the `IGNORE` query response, and this offers the basis for an explanation of the results: child-adult conversation generally requires an adult to provide answers to a child's questions, even if the question somehow deviates from the topic of the conversation.

5 ANNOTATION RELIABILITY

As we mentioned above, the annotation process was performed by the first author. In order to check the reliability of the classification process, inter- and intra-annotator studies were performed.

For the inter-annotator study, a sample of 100 randomly chosen examples of query – query responses (retrieved from all four corpora we utilised) was used. The distribution of the classes was in line with the distribution observed by the primary annotator: `CR`: 31 examples; `DP`: 32 examples; `MOTIV`: 11 examples; `NO ANSW`: 8 examples; `FORM`: 5 examples; `IND`: 7 examples; `IGNORE`: 6 examples.

All the examples were supplemented with a context. The guideline for annotators contained explanations of all the classes and examples of question-responses for each category. Also the `OTHER` category was included. The instruction was to annotate the query response to the first question in each example. The control sample was annotated by four annotators (three experienced linguists and a logician with moderate experience in corpus annotation).

The reliability of the annotation was evaluated using κ (Carletta 1996), established by using the R statistical software (R Core Team 2013; version 3.1.2) with the *irr* package (Gamer *et al.* 2012). The interpretation of the kappa values is based on that of Viera and Garrett 2005.

The Fleiss κ for all five annotators was 0.64 (i.e. substantial) with 51% agreement over 100 cases. The agreements between the main annotator and others were all substantial:

1. main and second annotator: $\kappa = 0.67$ with 73% agreement;
2. main and third annotator: $\kappa = 0.65$ with 72% agreement;
3. main and fourth annotator: $\kappa = 0.61$ with 69% agreement;
4. main and fifth annotator: $\kappa = 0.63$ with 70% agreement.

When it comes to detailed analysis of the annotation we start with the `OTHER` category. The annotators were given the option of using this category and, in fact, this option was not used frequently (from a sample of 100 cases): annotator 1: 0, annotator 2: 3, annotator 4: 6, annotator 4: 8, annotator 5: 0.

When we take a closer look at the disagreements between the main and the fourth annotator (the lowest agreement observed) what becomes clear is that the most problematic cases were `DP` vs. `IGNORE` (5 cases). This fact is quite surprising since these categories are rather distant. This suggests that the fourth annotator had misunderstood the category `IGNORE`.

An analysis of the cases involving the most disagreement suggests that these are not infrequently cases where genuine ambiguities exist given the intentional nature of many of the dialogical relations tying together queries. These naturally enough get exacerbated for annotators required to make decisions in a largely context independent manner.

Thus, (14) was annotated as `DP` by two annotators, as `IGNORE` by two annotators, and as `NO ANSW` by one annotator; in the context of adult/child interaction, `IGNORE` is possibly more likely – the child observing the same situation as the parent but ignoring politeness in trying to impose the issue momentarily captivating her own interest; at the same time a `DP` reading is potentially plausible given the plausibility of the assumption *What a fireman does with his axe depends on where his axe is*.

(14) between `DP` and `IGNORE`: Parent: what does a fireman do with his axe? Child: where's his axe is?

In a similar fashion, (15) was annotated as `DP` by three annotators and by two annotators as `NO ANSW`. Both are potentially plausible classifications: in a cooperative setting (e.g., a dinner enjoyed by a couple in a restaurant) `DP` is more appropriate (What Norrine will have as a starter depends on what Chris wants), whereas in a more adversarial setting the query response can simply be a means of avoiding the initial question.

(15) between `DP` and `NO ANSW`: Chris: What would you like to have start with? Norrine: What do you want?

In light of this issue, we hypothesize that a more satisfying account of annotator reliability for this task would involve developing an annotation model that accommodates ambiguity in annotation, as for instance in work on the basis of crowdsourced labels, as pioneered in Passonneau and Carpenter 2013. This constitutes work we hope to perform in the future.

For the intra-annotator study another control sample of 100 examples was randomly chosen from the data. The distribution of the classes was similar to that in the first control sample. In this case the agreement of the coding between the first annotation and that obtained in the study was substantial ($\kappa = 0.78$).

6 MODELING QUERY RESPONSE CATEGORIES IN KOS

6.1 *Dialogue gameboards, conversational rules, and dialogical relevance*

We offer a formal explication of the coherence that underlies the various different types of query responses within the framework of KoS (Ginzburg 2012). We offer here an analogy to formal syntax. When one discovers a class of constructions C in need of analysis, one means of showing that a given formalism F is adequate involves showing that F's weak (strong) generative capacity properly includes the string set (analysis trees, etc.) corresponding to C. Within dialogue similar desiderata exist, where constructions are replaced by pairs (or longer sequences) of coherent utterances.[7] We seek to show that KoS's notion

[7] An anonymous reviewer for this journal cautions us about the analogy between syntactic grammaticality and dialogue coherence. We agree that the analogy with syntax should not be exaggerated. There are differences. But the analogy between syntactic ungrammaticality and dialogical incoherence is not entirely far-fetched: if one says something incoherent, one could be adjudged to be linguistically incompetent, just as with ungrammaticality. With the latter one can use repair mechanisms to fix ungrammaticality ('I know who did Mary like, I mean who she liked.'), just as with the former (**A:** Who came yesterday? **B:** I'm having a coffee. **A:** Did you hear what I said? **B:** Oh sorry um yes, no I have no idea.). Of course, given the possibility of interpreting incoherence as intended irrelevance, one can often draw that as a possible inference, but grammaticality errors also potentially push us to expect repair, to view the other speaker as momentarily confused etc. The same reviewer points out that weak/strong generative capacity are not necessarily notions to be held as some kind of ideal for scientific explanation. Whatever one thinks of those notions, we think that they were, nonetheless, useful in stimulating syntactic research in the 60s to 80s, in trying to figure out which constructions stretch a given formalism to its limit (e.g., phrase structure grammars and cross-serial dependencies.). Similar considerations apply at present *mutatis mutandis* to formal dialogue theory. We return to this issue and how it relates to cross-theoretical comparison in Section 7.

of coherence properly includes the class of queries and their questions responses, a demonstration that to the best of our knowledge has not hitherto been attempted for any dialogue formalism.[8]

KoS is a framework for dialogue formulated using Type Theory with Records (TTR; Cooper 2005, 2012; Cooper and Ginzburg 2015). It provides formal underpinnings for the information state approach to dialogue management (Larsson and Traum 2003) and underlies dialogue systems such as GoDiS (Larsson 2002) and CLARIE (Purver 2006). On the approach developed in KoS, there is actually no single context. Instead, analysis is formulated at a level of information states, one per conversational participant. The type of such information states is given in (16a). We leave the structure of the private part unanalysed here, as with one exception none of our characterizations makes reference to this; for one approach to *private*, see Larsson 2002. The dialogue gameboard represents information that arises from publicized interactions. Its structure is given in (16b) – 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 being or have been grounded, *QUD* tracks the questions currently under discussion:[9]

(16) a. $\text{TotalInformationState (TIS)} \stackrel{\text{def}}{=} \begin{bmatrix} \text{dialoguegameboard : DGBType} \\ \text{private : Private} \end{bmatrix}$

b. $\text{DGBType} \stackrel{\text{def}}{=} \begin{bmatrix} \text{spkr: Ind} \\ \text{addr: Ind} \\ \text{utt-time : Time} \\ \text{c-utt : addressing(spkr,addr,utt-time)} \\ \text{Facts : Set(Prop)} \\ \text{Pending : list(LocProp)} \\ \text{Moves : list(LocProp)} \\ \text{QUD : poset(Question)} \end{bmatrix}$

[8] Ginzburg (2010, 2012) sketched such a characterization for the entire class of queries and their responses, though without a detailed corpus study.

[9] The motivation for *Pending* and the type Loc(utionary)Prop(osition) is explained in Section 6.4.

A dialogue gameboard $c1$ will be a record $r1$ such that (17a) holds; by definition this means that:[10] (i) the set of labels of $r1$ needs to be a superset of the set of labels of DGBType and (ii) for each judgement constituent of DGBType $l_k : T_k$, the value r1 gets for that label, denoted by $r1.l_k$, it is the case that $r1.l_k : T_k$. Thus, concretely in this case, $r1$ should have the make-up in (17b), and the constraints in (17c) need to be met:

(17) a. r1 : DGBType

b. $\begin{bmatrix} \text{spkr} = \text{A} \\ \text{addr} = \text{B} \\ \text{utt-time} = \text{t1} \\ \text{c-utt} = \text{p}_{utt(A,B,t1)} \\ \text{Facts} = \text{cg1} \\ \text{Pending} = \langle \text{p1},\ldots,\text{p}k \rangle \\ \text{Moves} = \langle \text{m1},\ldots,\text{m}k \rangle \\ \text{QUD} = \text{Q} \end{bmatrix}$

c. A: Ind, B: Ind, t1 : Time, $\text{p}_{utt(A,B,t1)}$: addressing(A,B,t1), cg1 : Set(Prop), $\langle$p1,…,p$k\rangle$: list(LocProp) $\langle$m1,…,m$k\rangle$: list(LocProp), Q : poset(Question)

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 subtypes of DGBType. We explain briefly how this allows one to capture the coherence of responses.

We start by specifying how a question becomes established as in the DGB. The rule in (18) says that given a question q and ASK(A,B,q) being the LatestMove, one can update QUD with q as the maximal element of QUD (henceforth, a *QUD-maximal* element or Max-QUD, the "discourse topic"):[11]

[10] For a more detailed discussion and exemplification, see Cooper and Ginzburg 2015, Section 2.2.

[11] Here, as in the rest of the paper, we make use of *manifest fields* (Coquand *et al.* 2003). A manifest field $[\ell = a{:}T]$ is a convenient notation for $[\ell{:}T_a]$ where

(18) Ask QUD-incrementation

$$\left[\begin{array}{lll}
\text{pre} & : & \left[\begin{array}{l}
\text{spkr : Ind} \\
\text{addr : Ind} \\
\text{utt-time : Time} \\
\text{c-utt : addressing(spkr,addr,utt-time)} \\
\text{Facts : Set(Prop)} \\
\text{Pending : list(LocProp)} \\
\text{q : Question} \\
\text{Moves } = \left\langle \text{Ask(spkr,addr,q),m0} \right\rangle : \\
\qquad\qquad\qquad\qquad\qquad \text{list(LocProp)} \\
\text{QUD : poset(Question)}
\end{array}\right] \\
\text{effects} & : & \left[\begin{array}{l}
\text{spkr } = \text{ pre.spkr : Ind} \\
\text{addr } = \text{ pre.addr : Ind} \\
\text{utt-time } = \text{ pre.utt-time : Time} \\
\text{c-utt : addressing(spkr,addr,utt-time)} \\
\text{Facts } = \text{ pre.Facts : Set(Prop)} \\
\text{Pending } = \text{ pre.Pending : list(LocProp)} \\
\text{Moves } = \text{pre.Moves : list(LocProp)} \\
\text{QUD } = \left\langle \text{pre.q,pre.QUD} \right\rangle : \text{ poset(Question)}
\end{array}\right]
\end{array}\right]$$

In order to avoid the prolixity exemplified in (18), the rules in this paper employ a number of abbreviatory conventions. First, instead of specifying the full value of the list Moves, we usually record merely its first member, which we call ‘LatestMove’. Second, the preconditions can be written as a *merge* of two record types $DGBType^{-} \wedge_{merge} PreCondSpec$, one of which $DGBType^{-}$ is a subtype of DGBType and therefore represents predictable information common to all conversational rules; *PreCondSpec* represents information specific to the preconditions of this particular interaction type. Similarly, the effects can be written as a merge of two record types $DGBType^{0} \wedge_{merge} ChangePrecondSpec$, where $DGBType^{0}$ is a supertype of the preconditions and *ChangePrecondSpec* represents those aspects

T_a is a singleton type whose only witness is a. Singleton types are introduced by the clauses in (18).
1. If $a : T$ then T_a is a type.
2. $b : T_a$ iff $b = a$.

of the preconditions that have changed.[12] So we can *abbreviate* (18) as (19b):

(19) a. $\left[\begin{array}{lll} \text{pre} & : & \text{PreCondSpec} \\ \text{effects} & : & \text{ChangePrecondSpec} \end{array}\right]$

b. Ask QUD-incrementation

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

We can exemplify how this rule works. Assume (20a) to be a record that satisfies the preconditions of the type (19b), in other words it is a record which is of the type assigned to 'pre' in (18) or in abbreviated form in (19b). Hence, it constitutes the appropriate context for Ask QUD-incrementation. The output of that rule is (20b):

(20) a. $\left[\begin{array}{lll} \text{spkr} & = & \text{A} \\ \text{c1} & = & \text{p1} \\ \text{addr} & = & \text{B} \\ \text{c2} & = & \text{p2} \\ \text{r} & = & \text{q0} \\ \text{LatestMove} & = & \text{Ask(A,B,q)} \\ \text{QUD} & = & \left\langle\right\rangle \\ \text{FACTS} & = & \text{cg1} \end{array}\right]$

b. $\left[\begin{array}{lll} \text{spkr} & = & \text{A} \\ \text{addr} & = & \text{B} \\ \text{r} & = & \text{q0} \\ \text{LatestMove} & = & \text{Ask(A,B,q0)} \\ \text{QUD} & = & \left\langle \text{q0} \right\rangle \\ \text{FACTS} & = & \text{cg1} \end{array}\right]$

We also assume an analogue of (19b) for assertion, given in (21). In an interactive setting A asserting p raises the issue p? for B – s/he

[12] This procedure is described in much more general terms using the operation of *asymmetric* merge by Cooper (2016), who shows the use of this operation for a wide range of semantic uses.

can then either decide to discuss this issue (as a consequence of the rule QSPEC introduced below as (24)) or accept it as positively resolved (as a consequence of the rule (22)):

(21) Assertion QUD-incrementation

$$\begin{bmatrix} \text{pre :} & \begin{bmatrix} \text{p :} & \text{Prop} \\ \text{LatestMove} = \text{Assertion(spkr,addr,p):} & \text{LocProp} \end{bmatrix} \\ \text{effects :} & \begin{bmatrix} \text{QUD} = \left\langle \text{p?,pre.QUD} \right\rangle\text{:} & \text{poset(Question)} \end{bmatrix} \end{bmatrix}$$

An obvious complement to QUD incrementation is a principle controlling QUD downdate. Since QUD consists of questions that are *unresolved* relative to FACTS, QUD downdate is formulated simultaneously with FACTS update: when p is added to FACTS, one needs to verify for all existing elements of QUD that they are not resolved by the new value of FACTS. This joint process of FACTS update / QUD downdate is formulated in (22): given an acceptance or confirmation of p by B, p can be unioned into FACTS, whereas QUD is modified by the function NonResolve. NonResolve is a function that maps a partially ordered set of questions *poset*(q) and a set of propositions P to a partially ordered set of questions *poset*$'(q)$ which is identical to *poset*(q) modulo those questions in *poset*(q) resolved by members of P.

(22) a. Fact Update/ QUD Downdate $\stackrel{\text{def}}{=}$

$$\begin{bmatrix} \text{pre} & : \begin{bmatrix} \text{p} & : & \text{Prop} \\ \text{LatestMove} = \text{Accept(spkr,addr,p)} & : & \text{LocProp} \\ \text{QUD} = \left\langle \text{p?,pre.QUD} \right\rangle & : & \text{poset(Question)} \end{bmatrix} \\ \text{effects} & : \begin{bmatrix} \text{FACTS} = \text{pre.FACTS} \cup \left\{ \text{p} \right\} & : & \text{Set(Prop)} \\ \text{QUD} = \text{NonResolve(pre.QUD,FACTS)} & : & \text{poset(Question)} \end{bmatrix} \end{bmatrix}$$

b. NonResolve $\stackrel{\text{def}}{=}$

$$\text{r} : \left(\begin{bmatrix} \text{B : Ind} \\ \text{P : set(Prop)} \\ \text{Q : poset(InfoStruc)} \end{bmatrix} \right) \begin{bmatrix} \text{Q}' : \text{poset(InfoStruc)} \\ \text{c1} : \text{Q}' \subset \text{r.Q} \\ \text{c2} : \forall q_0 \in \text{Q}' \neg \exists f \in P \\ \mathit{Resolve}(f, q_0.q) \end{bmatrix}$$

With this in hand, we now turn to explaining how dialogical relevance is handled in KoS. Pre-theoretically, relevance relates an utterance u to an information state I just in case there is a way to successfully update I with u. Ginzburg (2012) defines two notions of relevance, a simpler one at the level of moves, i.e. illocutionary contents of utterances, and a somewhat more complex one at the level of utterances. For expository simplicity, we restrict attention here to the former and refer the reader to Ginzburg 2010, 2012 for the more complex notion.

The basic concept introduced here is contextual *m(ove)-coherence* defined in (23a) as applying to m_1 and dgb_0 just in case there is a conversational rule c_1 which maps dgb_0 to dgb_1 and such that dgb_1's LatestMove value is m_1. **Pairwise M(ove)-Coherence**, defined in (23b), applies to a pair of moves m_1, m_2, if m_1 is **M-Coherent** relative to some DGB dgb_0 and there is a sequence of updates leading from LatestMove being m_1 to LatestMove being m_2. Finally, **Sequential M(ove)-Coherence**, defined in (23c), applies to a sequence of moves $m_1, \ldots, m_n$ just in case each successive pair of moves are **Pairwise M-Coherent**:

(23) a. **M(ove)-Coherence**: Given a set of conversational rules $\mathscr{C}$ and a dialogue gameboard dgb_0 : DGBType, a move m_1 : LocProp is $\mathbf{m(ove)}_{\mathscr{C}}{}^{dgb_0}$**-coherent** iff
(i) there exists $dgb_1 : DGBType, c_1 \in \mathscr{C}$ such that $c_1(dgb_0) = dgb_1$ and
(ii) $dgb_1.LatestMove = m_1$.

b. **Pairwise M(ove)-Coherence**: Given a set of conversational rules $\mathscr{C}$ two moves m_1, m_2 are $\mathbf{m(ove)}_{\mathscr{C}}$**-pairwise-coherent** iff there exists $dgb_0 : DGBType$ and $dgb_i, c_i, (1 \leq i \leq k-1, dgb_i : DGBType, c_i \in \mathscr{C})$ such that
(i) m1 is $\mathbf{m(ove)}_{\mathscr{C}}{}^{dgb_0}$**-coherent** and
(ii) $c_{i+1}(dgb_i) = dgb_{i+1}$ and $dgb_i.LatestMove = m_1$, whereas $dgb_k.LatestMove = m_2$.

c. **Sequential M(ove)-Coherence**: A sequence of moves m_1 ,…, m_n is $\mathbf{m}_{\mathscr{C}}$**-coherent** iff for any $1 \leq i$ m_i, m_{i+1} are $\mathbf{m}_{\mathscr{C}}$**-pairwise-coherent**.

6.2 *Question accepting responses*

6.2.1 The class DP

We start by characterizing the moves in which the responder B accepts question $q1$ as an issue to be resolved. The potential for DP responses is explicated on the basis of QSPEC, the conversational rule in (24a). This rule characterizes the contextual background of reactive queries and assertions. It specifies that if q is QUD-maximal, then subsequent to this either conversational participant may make a move constrained to be q-specific, conveying either a proposition p which is a partial answer to q (p is *about* q) or a question $q1$ on which q *depends*, as defined in (24c); one possible definition of *dependence* is given in (24d); intuitively the idea is that if q is dependent on $q1$, then once one knows an answer that resolves $q1$, some information about q (viz. a partial answer) becomes available. This originates in Ginzburg (2012), where formal characterizations of *aboutness* and *resolvedness* can be found.[13]

(24) a. QSPEC

$$\begin{bmatrix} \text{pre} : \begin{bmatrix} \text{QUD} = \left\langle \text{q}, \text{Q} \right\rangle : \text{poset(Question)} \end{bmatrix} \\ \text{effects} : \text{TurnUnderspec} \wedge_{merge} \\ \begin{bmatrix} \text{r} : \text{Question} \vee \text{Prop} \\ \text{R: IllocRel} \\ \text{LatestMove} = \text{R(spkr,addr,r)} : \text{LocProp} \\ \text{c1} : \text{Qspecific(r,q)} \end{bmatrix} \end{bmatrix}$$

b. q-specific utterance: an utterance whose content is either a proposition p About q or a question $q1$ on which q Depends

c. $q1$ depends on $q2$ iff any proposition p such that p resolves $q2$, also satisfies that for some r p entails r such that r is about $q1$.

Other characterizations of dependency are conceivable and could replace the one given here. For now, we illustrate how dependent responses emerge as relevant responses: in (25) A asks $q1$, responded to by B with a dependent question response $q2$. A answers $q2$, which gets accepted by B, leading to an answer to $q1$:

[13] We notate the underspecification of the turn holder as *TurnUnderspec*, an abbreviation for the following specification which gets unified together with the rest of the rule:

$$\begin{bmatrix} \text{PrevAud} = \left\{ \text{pre.spkr,pre.addr} \right\} & : & \text{Set(Ind)} \\ \text{spkr} & : & \text{Ind} \\ \text{c1} & : & \text{member(spkr, PrevAud)} \\ \text{addr} & : & \text{Ind} \\ \text{c2} & : & \text{member(addr, PrevAud)} \\ & & \wedge \text{ addr} \neq \text{spkr} \end{bmatrix}$$

(25) **A**(1)**:** Who should we invite for tomorrow?
B(2)**:** Who will agree to come?
A(3)**:** Helen and Jelle and Fran and maybe Sunil.
B(4)**:** (a) I see. (b) So, Jelle I think.
A(5)**:** OK.

(26)

Utt.	DGB Update (Conditions)	Rule
initial	MOVES $= \langle\rangle$	
	QUD $= \langle\rangle$	
	FACTS = cg1	
1	LatestMove := Ask(A,B,$q1$)	
	QUD : $= \langle q1\rangle$	Ask QUD-incrementation
2	LatestMove := Ask(B,A,$q2$)	QSPEC
	Influence($q2, q1$)	
	QUD : $= \langle q2, q1\rangle$	Assert QUD-incrementation
3	LatestMove := Assert(A,B,$p2$)	QSPEC
	(About($p2, q2$))	
	QUD := $\langle p2?, q2, q1\rangle$	Assert QUD-incrementation
4a	LatestMove := Accept(B,A,$p2$)	Accept
	FACTS := cg1 $\cup$ $\{p2\}$	
	QUD := $\langle q1\rangle$	Fact update/QUD downdate
4b	LatestMove := Assert(B,A,$p1$)	QSPEC
	(About($p1, q1$)	
	QUD := $\langle p1?, q0\rangle$	Assert QUD-incrementation
5	LatestMove := Accept(A,B,$p1$)	Accept
	FACTS := cg1 $\cup\{p1, p2\}$	
	QUD := $\langle q1\rangle$	Fact update/QUD downdate

6.2.2 The class IND

This class consists of query responses, where $q2$ is posed rhetorically, and which provide (indirectly) an answer to $q1$. In other words, $q2$ is posed in a context where an answer that resolves $q2$ can be assumed to be in FACTS – the repository of shared assumptions in the DGB, and, moreover, this answer entails a (resolving) answer to $q1$. Handling this class does not involve any additional conversational rules; it requires solely two independently needed additions to the setup described hitherto, a mechanism for rhetorical interpretation of questions and a means of accommodating *indirect* answers:

1. **Rhetorical interpretation of interrogatives**: a rhetorical use arises when an interrogative $q1$ is used in a context where the DGB contains a fact f that resolves $q1$. There are, in fact, two possible ways this can be satisfied: either the question has been discussed and a resolving answer provided; alternatively, certain answers are default values for such uses – a negative universal for wh-questions (‘Who cares?’, ‘Who knows?’), one of the polar values for polar-questions (‘Do I care?’, ‘Is the Pope Catholic?’). One possible treatment is proposed in Ginzburg 2012, §8.3.5: given a context in which a proposition p resolving a question q is presupposed, one postulates a root construction that assigns a clause denoting a question q the force of a reassertion of p, where p, a proposition resolving q, is a presupposition satisfied by the context.
2. **Indirect answers**: we need to allow q-specificity to include propositions that are *indirectly* about q. An explicit account of indirectness would take us too far afield here, but see e.g., Asher and Lascarides 1998 and Ginzburg 2012, §8.3.3–8.3.5 for discussion relating to questions in dialogue.

We exemplify how this works in (27): A asks the question $p0?$ (‘Is B’s job safe?’). B responds with a reassertion of a question $q1$ (‘Whose job is safe?’), which in this context reasserts the proposition $p1$ ‘No one’s job is safe.’, which in particular resolves the question $p0?$. This explains *inter alia* why there is no need for A to respond to B’s question.

(27) **A:** Is your job safe?
B: Whose job’s safe?

Utt.	DGB Update (Conditions)	Rule
initial	MOVES = $\langle\rangle$	
	QUD = $\langle\rangle$	
	FACTS = cg1	
1	LatestMove := Ask(A,B,$p0?$)	
	QUD : = $\langle p0? \rangle$	Ask QUD-incrementation
2	LatestMove := ReAssert(B,A,$p1$)	QSPEC
		Resolve($p1, q1$)

We mentioned previously a subclass of IND – query responses where $q2$ *presupposes* an answer that resolves $q1$. We do not offer a detailed analysis of this subclass here, but they could, for instance, be accommodated by a slight adjustment of q-specificity which licensed responses whose content semantically presupposed a proposition p about $q1$.

6.2.3 The class FORM

This class consists of query-responses addressing the issue of the *way* the answer to $q1$ should be given. The class FORM raises interesting issues since it seems to be the sole class whose coherence intrinsically involves reasoning by the responder about the original querier's unpublicized intentions. One possible explication would be in terms similar to the relation Q-Elab (Asher and Lascarides 2003). Perhaps the simplest way to do this in the current setting would be, following Larsson (2002), to widen the definition of q-specificity so that it is relative to an information state which provides a notion of the agent's plan, decomposed into a sequence of questions to be resolved:

(28) u is q-specific relative to an information state I: an utterance whose content is either a proposition p About q or a question $q1$ on which q Depends or a question q_1' which is a component of I.plan

One could try and collapse DP and FORM. One reason not to do this is precisely that the former does not intrinsically involve reasoning about intentions and so, in principle, its coherence should be easier to compute.

6.3 *Adversarial query responses*

Adversarial query responses are challenging for most semantic theories of questions, for reasons we discuss below. Common to all three classes is a lack of acceptance of $q1$ as an issue to be discussed. In MOTIV-type responses the need/desirability to discuss $q1$ is explicitly posed, in NO ANSW-type responses there is an implicature that $q1$ is of lesser importance/urgency than $q2$, whereas for IGNORE type responses there is an implicature that q1 as such will not be addressed.

One commonality between `MOTIV` and `NO ANSW` worth noting is that in both cases $q1$ actually needs to be added to QUD at the outset. One might think that a consequence of a responder's failure to accept q for discussion is that q will only resurface if explicitly reposed. There is evidence, however, that actually q remains in a conversational participant's QUD even when not initially adopted, as its very posing makes it temporarily DGB available. In (29), where move (2) could involve either a `MOTIV` query (2a), or a `NO ANSW` query (2b), the original question has definitely *not* been re-posed and yet B still has the option to address it, which s/he should be unable to do if it is not added to his gameboard before (29(2)).

(29) **A:** Who are you meeting next week?
B(2)**:** (2a) What's in it for you? / (2b) Who are *you* meeting next week?
A: I'm curious.
B: Aha.
A: Whatever.
B: Oh, OK, Jill.

We turn to a discussion of the coherence of each class, starting with `MOTIV` and `NO ANSW`, leaving `IGNORE` for later, given a certain additional complexity it embodies.

6.3.1 The class `MOTIV`

`MOTIV` utterances are an instance of metadiscursive interaction – interaction about what should or should not be discussed at a given point in a conversation, as exemplified by utterances such as (30):

(30) a. I don't know.
b. Do we need to talk about this now?
c. I don't wish to discuss this now.

A natural way to analyze such utterances is along the lines of a rule akin to QSPEC given in (24): q being MaxQUD gives (the responder) B the right to follow up with an utterance specific to the issue we could paraphrase informally as *?WishDiscuss(B,q)*.[14] Such a rule is formulated in (31), where the notation

[14] We are formulating this rule asymmetrically with respect to the interlocuters, in contrast to QSPEC, since A posing $q1$ means that A keeping the turn

$$\text{`QUD} = \left\langle \text{Max} = \left\{\text{?WishDiscuss(B,q1),q1}\right\}, \text{Q} \right\rangle\text{'}$$

indicates that both ?WishDiscuss(B,q1) and $q1$ are maximal in QUD, unordered with respect to each other. The motivation for this latter is the need to integrate $q1$ in context, as per (29) above.

(31) MetaDiscussing q1

$$\begin{bmatrix} \text{pre} : \begin{bmatrix} \text{QUD} = \langle q1, \text{Q}\rangle : \text{poset(Question)} \end{bmatrix} \\ \text{effects} : \begin{bmatrix} \text{spkr} = \text{pre.addr} : \text{Ind} \\ \text{addr} = \text{pre.spkr} : \text{Ind} \\ \text{r} : \text{Question} \vee \text{Prop} \\ \text{R: IllocRel} \\ \text{Moves} = \langle \text{R(spkr,addr,r)} \rangle \oplus \text{pre.Moves} : \text{list(LocProp)} \\ \text{c1} : \text{Qspecific(R(spkr,addr,r),?WishDiscuss(spkr,pre.MaxQUD))} \\ \text{QUD} = \left\langle \begin{matrix} \text{Max} = \{\text{?WishDiscuss(spkr,}q1\text{),}q1\}, \\ \text{Q} \end{matrix} \right\rangle : \text{poset(Question)} \end{bmatrix} \end{bmatrix}$$

In case information is accepted indicating negative resolution of ?WishDiscuss(B,$q1$), then $q1$ may be downdated from QUD. This involves a minor modification of the Fact Update/QUD Downdate rule (see (22) above).[15]

We exemplify (31) in two ways. First, with a variant of (29), where B's rejection of a question leads to the downdating of $q1$; then, with a very similar analysis of a MOTIV query response. does not wish to discuss $q1$, hence s/he accommodates ?WishDiscuss(B, $q1$)

and uttering (30b,c) would be somewhat incoherent; the status of (30a) as a follow up to $q1$ is somewhat different: in the commonest case, where a query is posed because the querier does not know the answer, (30a) is redundant and somewhat infelicitous. In cases where $q1$ is uttered in the spirit of 'Here is an interesting issue to discuss', (i) seems acceptable:

(i) *I* don't know.

Whether this should be taken to imply that (30a) and (30b,c) should be licensed by distinct mechanisms is an issue we will not try to resolve here.

[15] All that this involves is a modification of the function NonResolve which fixes the value of QUD after the fact update: in its new definition it maps a poset of questions *poset*(q) and a set of propositions P to a poset of questions *poset*′(q) which is identical to *poset*(q) modulo those questions in *poset*(q) resolved by

into QUD and offers an utterance concerning this issue. A accepts B's assertion, so using the new version of fact-update/qud-downdate $q1$ can be downdated and either conversationalist could introduce a new topic, as in (32):

(32) **A**(1)**:** Who are you meeting next week?
B(2)**:** No comment.
A(3)**:** I see.
A/B(4)**:** What are you doing tomorrow?

Utt.	DGB Update (Conditions)	Rule
initial	MOVES $= \langle\rangle$	
	QUD $= \langle\rangle$	
	FACTS $=$ cg1	
1	LatestMove $:=$ Ask(A,B,$q1$)	
	QUD $: = \langle q1\rangle$	Ask QUD-incrementation
2	LatestMove $:= \langle$ Assert(B,A,$p1$) $\rangle$	Discussing u?
	QUD $:= \langle$p1? $\succ$?WishDiscuss($q1$), $q1\rangle$	Assertion QUD-incrementation
3	LatestMove $:= \langle$ Assert(B,A,$p1$) $\rangle$	Accept
	QUD $:= \langle\rangle$	Fact update/QUD downdate
	FACTS $:=$ cg1 $\cup$ {p1}	

We suggest that a dialogue like (33) works in a similar way: A's answer to B's question (33(2)) can satisfy B, which will lead to the question ?WishDiscuss(B, q1) being positively resolved, enabling B to downdate it from her QUD and address the question (33(1)).

members of P, **as well as those questions q for whom ?WishDiscuss(q) is negatively resolved.**

$$\begin{bmatrix} \text{pre} : \begin{bmatrix} \text{p : Prop} \\ \text{LatestMove} = \text{Accept(spkr,addr,p)} \\ \text{QUD} = \left\langle \text{p?,pre.QUD} \right\rangle : \text{poset(Question)} \end{bmatrix} \\ \text{effects} : \begin{bmatrix} \text{FACTS} = \text{pre.FACTS} \cup \left\{\text{p}\right\} : \text{Set(Prop)} \\ \text{QUD} = \text{NonResolve(pre.QUD,FACTS).Q}' \\ : \text{Poset(Question)} \end{bmatrix} \end{bmatrix}$$

$$\text{NonResolve} \stackrel{\text{def}}{=} \text{r} : \left(\begin{bmatrix} \text{B : Ind} \\ \text{F : set(Prop)} \\ \text{Q : poset(Question)} \end{bmatrix} \right) \begin{bmatrix} \text{Q}' : \text{poset(InfoStruc)} \\ \text{c1} : \text{Q}' \subset \text{r.Q} \\ \text{c2} : \forall q_0 \in \text{Q}' \neg\exists f \in \text{F} \\ \mathit{Resolve}(f, q_0.q) \\ \vee \mathit{Resolve}(f, ?\mathit{WishDiscuss}(r.B, q_0.q)) \end{bmatrix}$$

B not being satisfied with A's answer is entirely similar to (32) *mutatis mutandis*:

(33) **A**(1)**:** Who are you meeting next week?
B(2)**:** Why?
A(3)**:** I need to know which refreshments to buy.

Utt.	DGB Update (Conditions)	Rule
initial	MOVES = $\langle\rangle$	
	QUD = $\langle\rangle$	
	FACTS = cg1	
1	LatestMove := Ask(A,B,$q1$)	
	QUD : = $\langle q1\rangle$	Ask QUD-incrementation
2	LatestMove := $\langle$ Ask(B,A,$q2$) $\rangle$	Discussing u?
	QUD := $\langle q2 \succ$?WishDiscuss(B,$q1$)$, q1\rangle$	Ask QUD-incrementation
3	LatestMove := Assert(A,B,$p1$)	QSPEC
	(About($p1, q2$))	
	QUD := $\langle p1? \succ q2 \succ$?WishDiscuss(B, $q1$)$, q1\rangle$	Assert QUD-incrementation
4a	LatestMove := Accept(B,A,$p1$)	Accept
	FACTS := cg1 $\cup$ {$p1$}	
	QUD := $\langle q0\rangle$	Fact update/QUD downdate

6.3.2 The class `NO ANSW`

`NO ANSW`-queries can be analysed in a fairly similar fashion. The main challenge such queries pose is to consider the coherence relation between $q1$ and $q2$. Unlike `IGNORE`, where it seems like there is little that need connect the two questions, save for some reference to the situation associated with $q1$, for `NO ANSW` the questions seem to need a fairly tight link. A tentative characterization of this link is the following: $q1$ and $q2$ are not *dependent* on each other, but instead there exists a third question, $q3$, such that $q3$ depends on $q1$ and $q3$ depends on $q2$. The rationale behind this characterization is that by responding with $q2$ B provides (a) an issue that is not unconnected with $q1$, but (b) it is informationally not subservient to $q1$. Hence, given that

$q3$ is (or can be accommodated to be) the general topic under discussion, $q2$ has an arguable case to being at least as discussion worthy as $q1$:[16]

(34) a. $q1$ = what do you (B) like? $q2$ = what do you (A) like? $q3$ = Who likes what?

b. $q1$ = Why should we buy that scanner? $q2$ = Why should we not buy that scanner? ; $q3$ = Should we buy that scanner?

Based on this, we define the relation of being *unifiably coherent*:

(35) Given $q1, q2$: Question $q1$ and $q2$ are unifiably coherent iff

1. Neither $q1$, nor $q2$ depend on the other: $\neg$Depend$(q2, q1)$ $\wedge$ $\neg$Depend$(q1, q2)$
2. There exists $q3$: Question which depends on both $q1$ and $q2$: Depend$(q3, q1)$ $\wedge$ Depend$(q3, q2)$

The potential for making such queries can be captured by the conversational rule in (36). Given that $q1$ is MaxQUD, the responder may respond with $q2$, assuming it to be unifiably coherent with $q1$. The immediate effect of this is to update QUD with the issue ?WishDiscuss(B,$q1$).

---

[16] An anonymous reviewer for this journal points out the following exchange as problematic for our taxonomy, suggesting that it is 'fully coherent given the sequel but the pair does not seem to fit any of the schemes':

(i) **A:** Are you coming on Friday?
B: Did you ever consider quarks?
A: No.
B: Well you should for your work and Friday there will be a lecture that is just right for you. I may be there myself.

Actually, we would suggest that this example would be classified as a `NO ANSW` by the annotation criteria we offer (since B views A's question as less important to consider than his and one could eliminate B's answer at the end without affecting coherence.). Nonetheless, it calls into question our formalized definition for `NO ANSW` in that it is not clear that the $q2$ and $q1$ are *unifiably coherent*. One might use this (constructed) example to argue for weakening the unifiable coherence clause. At the same time, it seems likely that B's response would initially be viewed as incoherent by A and this should be reflected by e.g., response time, frowning etc.

(36) Challenging q1

$$\left[\begin{array}{ll} \text{pre}: & \left[\text{QUD} = \left\langle q1, \text{Q}\right\rangle : \text{poset(Question)}\right] \\ \text{effects}: & \left[\begin{array}{l} \text{spkr} = \text{pre.addr} : \text{Ind} \\ \text{addr} = \text{pre.spkr} : \text{Ind} \\ q2 : \text{Question} \\ \text{Moves} = \left\langle \text{Ask(spkr,addr,}q2\text{)}\right\rangle \oplus \text{pre.Moves} : \text{list(LocProp)} \\ \text{c1} : \text{Unifiablycoherent(q1,}q2\text{)} \\ \text{QUD} = \left\langle \text{Max} = \left\{\begin{array}{l} \text{?WishDiscuss(B,}q1\text{),} \\ q1 \end{array}\right\}, \text{Q} \right\rangle : \\ \text{poset(Question)} \end{array}\right] \end{array}\right]$$

In (37)[17] A asks $q1$, B responds with $q2$ that unifies coherently with $q1$ via, for example, the issue $q3$ = ‘Should they wait?’. A responds to $q2$ and then B’s second utterance can be understood as addressing $q2$. If A accepts (4), $q2$ can be downdated and, consequently $q1$ and ?WishDiscuss(B,$q1$) as well – $q1$ has also been resolved, and hence ?WishDiscuss(B,$q1$) could be taken to be resolved as well.[18]

(37) **A**(1)**:** Why won’t they wait?
B(2)**:** Why should they?
A(3)**:** I waited.
B(4)**:** They have lives of their own.

6.4 *DGB divergence: Ignore and Clarification Requests*

Both clarification requests (CRs) and `IGNORE` type responses involve reasoning that requires reference to two DGBs. CRs arise due to a mismatch that occurs between what the speaker assumes her/his interlocutor’s linguistic/contextual knowledge is and what it actually is;

[17] Inspired by the BNC example:
Eddie: But it’s something, something in you, you have to rush don’t they? Why won’t they wait? Unknown: Why should they? Eddie: Why should they? Unknown: No, why should they? Eddie: I have Unknown: Take the rest of it Unknown: <*unclear*> Eddie: pleasure spending Unknown: <*unclear*> Unknown: No why, they’ve got lives of their own Eddie: Well Sally: let them live it, don’t want saving for the children, no, they don’t want nothing Eddie: Well Unknown: They’ve had far more than what we’ve ever had [KCF, 3584–3596].

[18] A general principle linking the downdating of ?WishDiscuss(B,q0) once $q0$ has been downdated should be introduced, though we will not do so here.

consequently, in the immediate aftermath of such an utterance – before the mismatch becomes manifest, the speaker updates her/his IS with the query s/he posed and the addressee updates hers/his with the clarification question s/he calculated.

Similarly, in the case of `IGNORE`s the initial speaker updates their information state with the query s/he posed and, ignoring this, the addressee updates hers/his with the situationally relevant question s/he has decided to pose.

6.4.1 Clarification Requests

We start by discussing CRs since they have been studied in great detail, see Ginzburg and Cooper 2004; Schlangen 2004; Purver 2006; Ginzburg *et al.* 2014; we will summarize briefly the most detailed account we are aware of, that provided in Ginzburg 2012. This will provide tools enabling us to analyse `IGNORE`-type responses.

Integrating clarification interaction into the DGB involves two modifications to the representations we have been using so far. One minor modification, drawing on an early insight of Conversation Analysis (Schegloff 2007), is that repair can involve 'putting aside' an utterance for a while, a while during which the utterance is repaired. That in itself can be effected without further ado by adding further structure to the DGB, specifically the field we call *PENDING*. 'Putting the utterance aside' raises the issue of *what is it that we are 'putting aside'*. In other words, how do we represent the utterance? The requisite information needs to be such that it enables the original speaker to interpret and recognize the coherence of the range of possible clarification queries that the original addressee might make. Ginzburg (2012) offers detailed arguments on this issue, including considerations of the phonological/syntactic parallelism exhibited between CRs and their antecedents, and the existence of CRs whose function is to request repetition of (parts of) an utterance. Taken together with the obvious need for PENDING to include values for the contextual parameters specified by the utterance type, Ginzburg concludes that the type of PENDING combines tokens of the utterance, its parts, and of the constituents of the content with the utterance type associated with the utterance. An entity that fits this specification is the *locutionary proposition* defined by the utterance. A locutionary proposition is a proposition whose situational component is an utterance situation, typed as in (38a), and will have the form in (38b):

(38) a. $LocProp \stackrel{\text{def}}{=} \begin{bmatrix} \text{sit} & : & \mathit{Sign} \\ \text{sit-type} & : & \mathit{RecType} \end{bmatrix}$

b. $\begin{bmatrix} \text{sit} = \text{u} \\ \text{sit-type} = T_u \end{bmatrix}$

Here T_u is a grammatical type for classifying u that emerges during the process of parsing u. It can be identified with a *sign* in the sense of Head-driven Phrase Structure Grammar (HPSG) (Pollard and Sag 1994).

How then can one characterize the relevance of CRs in this setup? Corpus studies of CRs (Purver *et al.* 2001; Rodriguez and Schlangen 2004; Rieser and Moore 2005) indicate that the subject matter of CRs is, in practice, restricted to three classes: CRs requesting repetition, CRs requesting confirmation, and CRs which query the intended content of a sub-utterance. This means that the potential for CRs can be modelled in terms of a small number of schemas (*Clarification Context Update Rules* (CCURs)) of the form: "if u is the maximal element of PENDING (*MaxPENDING*) and $u0$ is a constituent of u, add the clarification question $\text{CQ}^i(u0)$ into QUD.", where '$\text{CQ}^i(u0)$' is one of the three types of clarification question (repetition, confirmation, intended content) specified with respect to $u0$.

(39) is a simplified formulation of one CCUR, Parameter identification, which allows B to raise the following issue about A's sub-utterance $u0$: *what did A mean by u0?*:

(39) Parameter identification:

$$\begin{bmatrix} \text{pre} & : & \begin{bmatrix} \text{Spkr : Ind} \\ \text{MaxPENDING : LocProp} \\ \text{u0} \in \text{MaxPENDING.sit.constits} \end{bmatrix} \\ \text{effects} & : & \begin{bmatrix} \text{MaxQUD} = \lambda x \text{Mean(A,u0,x) : Question} \\ \text{LatestMove : LocProp} \\ \text{c1: CoPropositional(LatestMove.cont,MaxQUD)} \end{bmatrix} \end{bmatrix}$$

Here *CoPropositionality* for two questions means that, modulo their domain, the questions involve similar answers: for instance

'Whether Bo left', 'Who left', and 'Which student left' (assuming Bo is a student.) are all co-propositional. More generally, the definition is given in (40):

(40) Two utterances u_0 and u_1 are *co-propositional* iff the questions $q0$ and $q1$ they contribute to QUD are co-propositional.

a. qud-contrib(m0.cont) is m0.cont if m0.cont : Question
b. qud-contrib(m0.cont) is ?m0.cont if m0.cont : Prop[19]
c. $q0$ and $q1$ are co-propositional iff there exists a record r such that $q_0(r) = q_1(r)$.

Parameter Identification, as given in (39), underpins CRs such as (41b–41c) as follow-ups to (41a). Corrections can also be dealt with, as in (41d), since they address the issue of what A meant by u.

(41) a. **A:** Is Bo here?
b. **B:** Who do you mean 'Bo'?
c. **B:** Bo? (= Who is 'Bo'?)
d. **B:** You mean Jo.

To exemplify our account of how CRs get integrated in context, we exemplify in Figure 2 how the same input leads to distinct outputs on the "public level" of information states. In this case, it arises due to differential ability to anchor the contextual parameters. The utterance $u0$ has three sub-utterances, $u1$, $u2$, $u3$, given in Figure 2 with their approximate pronunciations. A can ground her/his own utterance since s/he knows the values of the contextual parameters, which we assume here for simplicity include the speaker and the referent of the sub-utterance *Bo*. This means that the locutionary proposition associated with $u0$ – the proposition whose situational value is a record that arises by unioning $u0$ with the witnesses for the contextual parameters and whose type is given in Figure 2 – is true. This enables the "canonical" illocutionary update to be performed: the issue *whether b left* becomes the maximal element of QUD. In contrast, assume that B lacks a witness for the referent of *Bo*. As a result, the locutionary proposition associated with $u0$ which B can construct is not true. Given this, B uses the CCUR *parameter identification* to build a context appropriate for a clarification request: B increments QUD with the issue

[19] Recall from the assertion protocol that asserting p introduces p? into QUD.

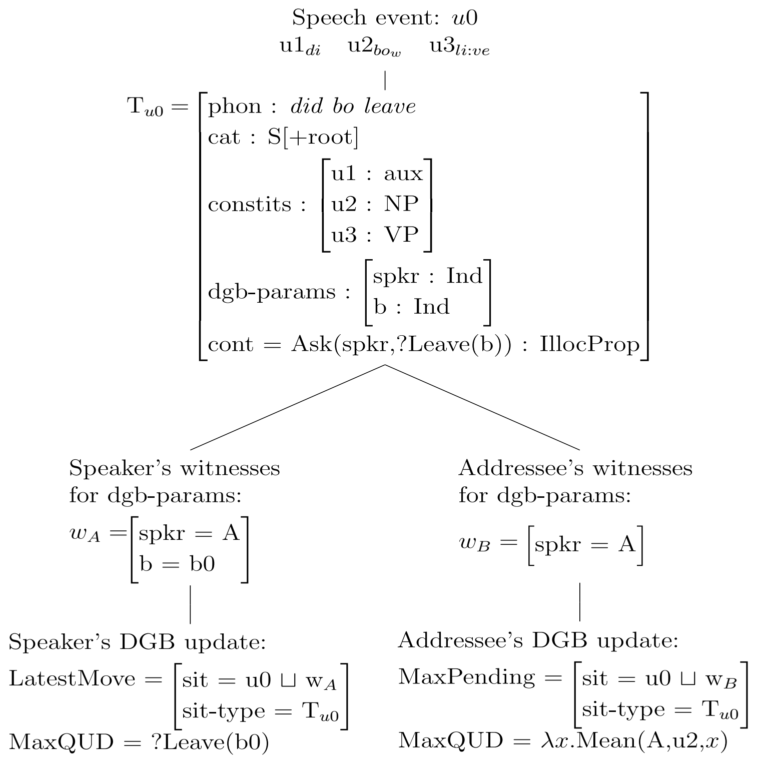


Figure 2: A single utterance giving rise to distinct updates of the DGB for distinct participants

λxMean(A,u2,x), and the locutionary proposition associated with $u0$ that B has constructed remains in PENDING.

6.4.2 The class IGNORE

The final class we consider is that of IGNORE-type responses. Such responses implicate that $q1$ will not be addressed, somewhat analogously to the classic Gricean floutings of relevance (A: Bob is an embarrassment B: It's very hot in here). Nonetheless, the effect such responses have is different from Gricean floutings, since these responses are situationally relevant, which appears to minimize significantly the potential impoliteness associated with ignoring $q1$. We think the difference between these two cases should be experimentally testable (e.g., response times for Gricean floutings should be significantly larger than for IGNOREs).

The conversational rule we propose allows the potential for $q2$ *and* captures the implicature concerning $q1$ being ignored. The formulation of such a rule presupposes a notion of *relevance* between the content of an utterance ($q2$) and the current context. We assume here the notion of relevance we mentioned in Section 6.1 and define *irrelevance* as *failure of relevance*: for an utterance u being IrRelevant to an information state I amounts to: there is *no way* to successfully update I with u. At the same time we assume that $q2$ being *situationally* relevant means that the open proposition component of $q2$ is of the form $p2(\ldots a \ldots)$, with a being in the situation which concerns $q1$.

This involves positing a conversational rule along the lines of (42) – given that (the content of) MaxPENDING – the most recent utterance, as yet ungrounded, hence maximal in PENDING – is *irrelevant* to the DGB but situationally relevant to $q2$, one can make MaxPENDING into LatestMove while updating Facts with the fact that the speaker of MaxPENDING does not wish to discuss MaxQUD:

(42) Ignoring questions

$$\begin{bmatrix} \text{pre} : & \begin{bmatrix} \text{a : IND} \\ \text{s1 : SIT} \\ \text{q1} = (\text{G})\begin{bmatrix} \text{sit} = \text{s1} \\ \text{sit-type} = \text{T} \end{bmatrix} : \text{Question} \\ \text{q2} = (\text{G1})\begin{bmatrix} \text{sit} = \text{s} \\ \text{sit-type} = \begin{bmatrix} \text{c : p2(a)} \end{bmatrix} \end{bmatrix} : \text{Question} \\ \text{In(s1,a)} \\ \text{dgb} = \begin{bmatrix} \text{MaxQUD} = \text{q1 : Question} \\ \text{MaxPENDING}^{content} = \text{q2 : Question} \end{bmatrix} : \text{DGBType} \\ \text{c: IrRelevant(MaxPENDING}^{content}\text{,dgb)} \end{bmatrix} \\ \text{effects} : & \begin{bmatrix} \text{LatestMove} = \text{pre.MaxPENDING : LocProp} \\ \text{Facts} = \text{pre.Facts} \cup \\ \left\{ \neg\ \text{WishDiscuss(pre.spkr,pre.MaxQUD)} \right\}. \end{bmatrix} \end{bmatrix}$$

Note that this does not make the *unwillingness to discuss* the *content* of the offending utterance; it is merely an inference. Still this inference will allow MaxQUD to be downdated, via fact update/question downdate, as was discussed with respect to MOTIV moves and the rule MetaDiscussing q1. We exemplify this with respect to (43).

(43) **A:** Is there just one car there?
B: Why is there no parking there?

As we noted earlier, given the contextual mismatch involved, in order to describe such dialogues one needs to consider the dialogue on the basis of two distinct DGBs. One possible evolution of A's DGB is this: A utters $q1$, which becomes MaxQUD; s/he then encounters B's response; A applies the rule *Ignoring questions*, which leads to $q1$'s downdate, $q2$ becomes MaxQUD.

(44)

Utt.	DGB Update (Conditions)	Rule
initial	MOVES = ⟨⟩	
	QUD = ⟨⟩	
	FACTS = cg1	
1	LatestMove := Ask(A,B,q1)	
	QUD : = ⟨$q1$⟩	Ask QUD-incrementation
2	LatestMove := ⟨ Ask(B,A,q2) ⟩	
	FACTS := FACTS ∪ ¬WishDiscuss(B,q1)	Ignoring questions
	QUD := ⟨ ⟩	FACTS update/QUD downdate
	QUD := ⟨ q2 ⟩	Ask QUD-incrementation

To the extent B wishes to ignore A's utterance, we do not need any additional machinery, save for a general principle needed in any case for a variety of other not necessarily linguistic events (e.g., in case one of the participants A burps, spits, or farts) – pretense that an event was not perceived. Assuming this, a possible evolution of B's DGB is as in (45): B pretends that A's utterance $u1$ did not take place, s/he utters $q2$, which relates to the situation A and B are jointly perceiving; $q2$ becomes MaxQUD:

(45)

Utt.	DGB Update (Conditions)	Rule
initial	MOVES = ⟨⟩	
	QUD = ⟨⟩	
	FACTS = cg1	
1	LatestMove := ⟨ Ask(B,A,q2) ⟩	
	QUD := ⟨ q2 ⟩	Ask QUD-incrementation

6.5 *Summary*

In this section we have shown how to characterize the relevance of the range of possible query responses $q2$ to an initial query $q1$ using DGB-based dynamics. The relevance of dependent questions is characterized in terms of QUD and the dependence relation, a relation defined on pairs of questions; `IND` uses the same contextual setup (plus mechanisms independently needed for accommodating rhetorical uses of interrogatives and indirect/presupposed answers); accommodating `FORM` involves reasoning similar to `DP`, but requires making reference to the issues constituting an interlocutor's plan; `MOTIV` and `NO ANSW` involve postulating additional conversational rules that make reference to the issue of whether $q2$'s speaker wishes to discuss $q1$, leaving this and $q1$ as issues simultaneously under discussion, hence this makes crucial use of QUD being a partially ordered set; `NO ANSW` also involves computing an additional coherence relation 'unifiable coherence' that needs to relate $q1$ and $q2$; clarification requests and `IGNORE` both require making reference to distinct DGBs for the two participants, make use of an additional buffer for ungrounded utterances, PENDING, and involve coherence relations defined at the level of utterances, not merely $q1$ and $q2$. The pre–theoretical complexity associated with each class is summarized in Table 9.

Table 9: Increasing complexity of reasoning needed to accommodate query responses

Query response type	Information state complexity
`DP`, `IND`	QUD, dependence relation
`FORM`	QUD, parametrised dep. relation
`MOTIV`	QUD as poset
`NO ANSW`	unif-coh relation, QUD as poset
`CR`, `IGNORE`	QUD, PENDING, DGB split non-semantic coherence

7 CONCLUSIONS

The article provides the first comprehensive, empirically based study of query responses to queries. One 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 designed our taxonomy based on data from the BNC since it is a general corpus with a variety of domains and genres, but have also shown that our classification works well in a number of more specific genres and domains, which display quite different distributions of query responses. We have proposed qualitative, domain-specific explanations for the variation displayed by these distributions.

On the theoretical side, we have provided a comprehensive, information state dynamics-based characterisation of the relevance of the entire range of query response types. Our account uses the KoS framework for representing dialogue information states and its component of information arising from publicized interaction, the dialogue game board (DGB). This enables us to offer a pre-theoretical sketch of the expressive complexity of the different classes of query response types, ranging from dependent questions and `IND`, which, assuming a semantic relation of question dependence, can be accommodated in a fairly vanilla query/response setup, through `MOTIV` and `NO ANSW`, which intrinsically require the dynamic question repository QUD to be a partially-ordered set, through `IGNORE` and clarification requests, which require distinct information DGBs for the two participants, make use of an additional buffer for ungrounded utterances, PENDING, and involve coherence relations defined at the level of utterances, not merely $q1$ and $q2$.

What are the more general theoretical implications of this characterization? We believe that it offers concrete desiderata for semantic theories, more specifically for the nature of conversational context. We offer brief remarks relative to frameworks that have put forward theories of question responses, as discussed in Section 1.

Some account of question dependence can be developed by any theory of questions which supplies notions of exhaustive and partial answerhood, though it is clear that providing a more detailed empirical and theoretical account of this notion than we have given here is an important task.

Relations like `MOTIV` and `NO ANSW` require structure within context since they need to maintain several questions simultaneously accessible to the participants. This constitutes a challenge for views of contexts in terms of *stacks*. Such a view has been made prominent in the view of QUD due to Roberts (1996). It can also be found, for instance, in the discourse model of Farkas and Roelofsen (2011), where a discourse context X is identified as a pair $\langle M, T \rangle$, where M is a Kripke model and T is a stack of sentences, those sentences that have been uttered so far.

The problem for stacks can be defused by adopting a distinct structure, for instance a partial order. Nonetheless, for these accounts and most other existing views of context, context is an entity shared by the conversational participants. This was also the case for the view of discourse structure in earlier work in SDRT (e.g., Asher and Lascarides 1998, 2003). In more recent work (e.g., Lascarides and Asher 2009), SDRT adopts a view advocated in KoS and also in the framework of PTT (Poesio and Traum 1998) that associates a distinct contextual entity with each conversational participant.

Given this, it seems that a framework like SDRT has potential for developing an account of question relations like IGNORE and CR which require context to 'diverge' across participants. There is one important caveat – we have argued that the notion of relevance that underpins both these question relations must make reference to non-semantic information. By contrast, in SDRT the semantics/pragmatics interface has no access to linguistic form, but only to a partial description of the content that is derived from linguistic form. This has been argued to be necessary to ensure the decidability of SDRT's glue logic (see e.g., Asher and Lascarides 2003, p. 77).

In closing, we note two questions raised by our account. The coherence follows in some cases on the basis of quite general conversational rules (e.g., QSPEC and MetaDiscussing q1) and in other cases on the basis of rather specific – though domain-independent – rules (e.g. Ignoring questions). An obvious theoretical issue is whether one can attain similar coverage on the basis of more "general" rules allied with some other very general pragmatic principles. A converse question is whether investigation of specific genres will lead to the need for genre-specific conversational rules for certain classes of question relations.

ACKNOWLEDGEMENTS

This is a much extended version of the article 'A corpus-based taxonomy of question responses' presented at the 10th International Conference on Computational Semantics (IWCS), Potsdam, March 2013 and subsequently presented in a extended version at the *Questions in Discourse* workshop in Stuttgart in May 2014.

We would also like to give our thanks to Dorota Leszczyńska-Jasion, Katarzyna Paluszkiewicz, Mariusz Urbański, Andrzej Wiśniewski, to three anonymous reviewers for the *Journal of Language Modelling,* to Elżbieta Hajnicz, to Adam Przepiórkowski, and to Carmela Chateau, for their help and insightful comments on this article.

We acknowledge support by the French Investissements d'Avenir-Labex EFL program (ANR-10-LABX-0083) and by the Disfluences, Exclamations, and Laughter in Dialogue (DUEL) project within the Projets Franco-Allemand en sciences humaines et sociales of the Agence Nationale de Recherche (ANR) and the Deutsche Forschunggemeinschaft (DFG); as well as by the Iuventus Plus grant (IP2011-031-771) and by the funds of the National Science Centre, Poland (DEC-2012/04/A/HS1/00715).

REFERENCES

Nicholas ASHER and Alex LASCARIDES (1998), Questions in dialogue, *Linguistics and Philosophy*, 21(3):237–309.

Nicholas ASHER and Alex LASCARIDES (2003), *Logics of conversation*, Cambridge University Press, Cambridge.

Jean CARLETTA (1996), Assessing agreement on classification task: The kappa statistic, *Computational Linguistics*, 22(2):249–254.

Lauri CARLSON (1983), *Dialogue games*, Synthese Language Library, D. Reidel, Dordrecht.

Robin COOPER (2005), Austinian truth in Martin-Löf Type Theory, *Research on Language and Computation*, 3(4):333–362.

Robin COOPER (2012), Type Theory and semantics in flux, in Ruth KEMPSON, Nicholas ASHER, and Tim FERNANDO, editors, *Handbook of the Philosophy of Science*, volume 14: Philosophy of Linguistics, pp. 271–323, Elsevier, Amsterdam.

Robin COOPER (2016), Type Theory and language: From perception to linguistic communication, `https://sites.google.com/site/typetheorywithrecords/drafts/ttl161130.pdf?attredirects=0`, book draft (access 20.03.2017).

Robin COOPER and Jonathan GINZBURG (2015), Type Theory with Records for NL semantics, in Chris FOX and Shalom LAPPIN, editors, *Handbook of Contemporary Semantic Theory, Second Edition*, pp. 375–407, Blackwell, Oxford,

doi:10.1002/9781118882139.ch12,
`http://dx.doi.org/10.1002/9781118882139.ch12`.

Thierry COQUAND, Randy POLLACK, and Makoto TAKEYAMA (2003), A logical framework with dependent types, *Fundamenta Informaticae*, 20:1–21.

Donka FARKAS and Floris ROELOFSEN (2011), Polarity particles in an inquisitive discourse model, `https://www.illc.uva.nl/inquisitivesemantics/papers/publications`, access 20.03.2017. Manuscript, University of California at Santa Cruz and ILLC, University of Amsterdam.

Matthias GAMER, Jim LEMON, and Ian Fellows Puspendra SINGH (2012), *irr: Various coefficients of interrater reliability and agreement*, `http://CRAN.R-project.org/package=irr`, acess 20.03.2017, R package version 0.84.

Jonathan GINZBURG (1995), Resolving questions, I, *Linguistics and Philosophy*, 18:459–527.

Jonathan GINZBURG (2010), Relevance for dialogue, in Paweł ŁUPKOWSKI and Matthew PURVER, editors, *Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue*, Polish Society for Cognitive Science, Poznań, iSBN 978-83-930915-0-8.

Jonathan GINZBURG (2012), *The interactive stance: Meaning for conversation*, Oxford University Press, Oxford.

Jonathan GINZBURG and Robin COOPER (2004), Clarification ellipsis, and the nature of contextual updates, *Linguistics and Philosophy*, 27(3):297–366.

Jonathan GINZBURG, Raquel FERNÁNDEZ, and David SCHLANGEN (2014), Disfluencies as intra-utterance dialogue moves, *Semantics and Pragmatics*, 7(9):1–64.

Arthur C. GRAESSER, Natalie K. PERSON, and John D. HUBER (1992), Mechanisms that generate questions, in Thomas W. LAUER, Eileen PEACOCK, and Arthur C. GRAESSER, editors, *Questions and information systems*, pp. 167–187, Lawrence Erlbaum Associates, Hillsdale.

Jeroen GROENENDIJK (2009), Inquisitive semantics: Two possibilities for disjunction, in Peter BOSCH, David GABELAIA, and Jérôme LANG, editors, *Logic, Language, and Computation*, volume 5422 of *Lecture Notes in Computer Science*, pp. 80–94, Springer Berlin / Heidelberg.

Jeroen GROENENDIJK and Floris ROELOFSEN (2011), Compliance, in Alain LECOMTE and Samuel TRONÇON, editors, *Ludics, Dialogue and Interaction*, pp. 161–173, Springer-Verlag, Berlin Heidelberg.

Jacqueline C. KOWTKO and Patti J. PRICE (1989), Data collection and analysis in the air travel planning domain, in *Proceedings of the Workshop on Speech and*

Natural Language, HLT ’89, pp. 119–125, Association for Computational Linguistics, Stroudsburg, PA, USA, ISBN 1-55860-112-0, doi:10.3115/1075434.1075455, `http://dx.doi.org/10.3115/1075434.1075455`.

Jan Van KUPPEVELT (1995), Discourse structure, topicality and questioning, *Journal of Linguistics*, 31:109–147.

Staffan LARSSON (2002), *Issue based dialogue management*, Ph.D. thesis, Gothenburg University, `http://www.ling.gu.se/~/sl/Thesis/thesis.pdf`, access 20.03.2017.

Staffan LARSSON and David TRAUM (2003), The information state approach to dialogue management, in Jan VAN KUPPEVELT and Ronnie SMITH, editors, *Advances in Discourse and Dialogue*, Kluwer.

Alex LASCARIDES and Nicholas ASHER (2009), Agreement, disputes and commitments in dialogue, *Journal of Semantics*, 26(2):109–158.

Brian MACWHINNEY (2000), *The CHILDES project: Tools for analyzing talk*, Lawrence Erlbaum Associates, Mahwah, NJ, third edition.

Rebecca J PASSONNEAU and Bob CARPENTER (2013), The benefits of a model of annotation, in *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pp. 187–195, Citeseer.

Michal PELIŠ and Ondrej MAJER (2010), Logic of questions from the viewpoint of dynamic epistemic logic, in M. PELIŠ, editor, *The Logica Yearbook 2009*, pp. 157–172, College Publications, London.

Massimo POESIO and David TRAUM (1998), Towards an axiomatization of dialogue acts, in J. HULSTIJN and A. NIJHOLT, editors, *Proceedings of TwenDial 98, 13th Twente workshop on Language Technology*, pp. 207–221, Twente University, Twente.

Carl POLLARD and Ivan A. SAG (1994), *Head-driven Phrase Structure Grammar*, University of Chicago Press and CSLI, Chicago.

Matthew PURVER (2001), SCoRE: A tool for searching the BNC, Technical Report TR-01-07, Department of Computer Science, King’s College London, `ftp://ftp.dcs.kcl.ac.uk/pub/tech-reports/tr01-07.ps.gz`.

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

Matthew PURVER, Jonathan GINZBURG, and Patrick HEALEY (2001), On the means for clarification in dialogue, in *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue*, pp. 116–125, Association for Computational Linguistics, Aalborg, Denmark, `http://www.dcs.qmul.ac.uk/~mpurver/papers/pgh01sigdial.pdf`.

R CORE TEAM (2013), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/, acess 20.03.2017.

Verena RIESER and Joanna MOORE (2005), Implications for generating clarification requests in task-oriented dialogues, in *Proceedings of the 43rd Meeting of the Association for Computational Linguistics*, Michigan.

Craige ROBERTS (1996), Information structure in discourse: Towards an integrated formal theory of pragmatics, *Working Papers in Linguistics-Ohio State University Department of Linguistics*, pp. 91–136, reprinted in Semantics and Pragmatics, 2012.

Kepa RODRIGUEZ and David SCHLANGEN (2004), Form, intonation and function of clarification requests in German task-oriented spoken dialogues, in Jonathan GINZBURG and Enric VALLDUVÍ, editors, *Proceedings of Catalog'04, The 8th Workshop on the Semantics and Pragmatics of Dialogue*, Universitat Pompeu Fabra, Barcelona.

Carolyn P. ROSÉ, Barbara Di EUGENIO, and Johanna D. MOORE (1999), A dialogue-based tutoring system for basic electricity and electronics, in Susanne P. LAJOIE and Martial VIVET, editors, *Artificial intelligence in education*, pp. 759–761, IOS, Amsterdam.

Emanuel SCHEGLOFF (2007), *Sequence organization in interaction*, Cambridge University Press, Cambridge.

David SCHLANGEN (2004), Causes and strategies for requesting clarification in dialogue, in *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue*, pp. 136–143.

Anthony J. VIERA and Joanne M. GARRETT (2005), Understanding interobserver agreement: The kappa statistic, *Family Medicine*, 37(5):360–363.

Andrzej WIŚNIEWSKI (1995), *The posing of questions: Logical foundations of erotetic inferences*, Kluwer AP, Dordrecht, Boston, London.

Andrzej WIŚNIEWSKI (2003), Erotetic search scenarios, *Synthese*, 134:389–427.

Andrzej WIŚNIEWSKI (2013), *Questions, inferences and scenarios*, College Publications, London.