



CLARIE: Handling Clarification Requests in a Dialogue System^{*}

MATTHEW PURVER (mpurver@stanford.edu)

Center for the Study of Language and Information
Stanford University, Stanford, CA 94305, USA

Abstract. This paper sets out a approach to *clarification requests* (CRs) general enough to cover all the major forms found in corpus data and specific enough to analyse the questions they ask about individual words and phrases. Its main features are a view of utterances as contextual abstracts with a radically abstracted semantic representation, and a view of CRs as standard utterances asking standard questions, but showing a particular kind of contextual dependence. It shows how it can be implemented computationally within a prototype text-based dialogue system, CLARIE, allowing it not only to generate CRs to clarify unknown reference and learn new words, but also to interpret and respond to user CRs, with both capabilities integrated within the standard dialogue processes and governed by empirical evidence.

Key words: clarification requests, contextual dependence, grounding, information-state-based dialogue

1. Introduction

Clarification requests (CRs), questions about a previous (sub-)utterance’s content or form, are common in dialogue (3–6% of human-human dialogue turns according to two recent corpus studies (Purver et al., 2003a, hereafter PGH) and (Rodríguez and Schlangen, 2004, hereafter R&S)) but have often not been paid a great deal of theoretical or implementational attention. Computational dialogue system designers have always recognised the requirement to signal inability to understand a user turn (or inability to do so to a reasonable degree of confidence), and systems are therefore usually able to produce outputs like “*I did not understand what you said. Please rephrase*” or “*You want to go to Paris, is that right?*” (Larsson, 2002), that

^{*} This work was carried out as part of the ROSSINI project (EPSRC grant GR/R04942/01). Thanks are due to the reviewers of this journal and the *Catalog 2004* workshop at which a preliminary version appeared; Kepa Rodríguez and David Schlangen for discussion of their corpus data; Staffan Larsson for assistance with the GoDiS system; David Traum, Patrick Healey, Raquel Fernández and Alex Gruenstein for several useful discussions; and most of all Jonathan Ginzburg for enormous help and support as the supervisor of my thesis, of which this paper summarizes and extends chapter 6.

reject or check entire utterances. However, they are not usually able to clarify problems in a finer-grained way. But as systems start dealing with more complex tasks and domains, and thus requiring more detailed and complete meaning representations, the scope for various forms of misunderstanding increases: new out-of-vocabulary words, ambiguous words and sentences, reference resolution problems. As argued by (Gabsdil, 2003), there is then an advantage to being able to clarify at the individual word or phrase level, making the user aware of the specific source of the problem.

It is also very unusual for systems to be able to understand and respond when the *user* asks a CR. Designers (very sensibly) try to avoid user CRs by making system prompts as clear and informative as possible, and sometimes training users with those prompts. However, as systems start to deal with complex tasks, wider domains and wider audiences, and perhaps as they become more human-like and begin to be treated more like humans (Reeves and Nass, 1996), it seems inevitable that they will have to deal with users asking CRs at some point. When they do, it will be important to respond correctly, as in this imagined example:

- | | |
|-----|---|
| (1) | System: There are flights available via Paris or Amsterdam.
User: Paris?
System: (a) Yes, Paris.
(b) Paris, France.
(c) Paris is the quickest route, although Amsterdam is the cheapest.
(d) OK. Your ticket via Paris will be posted to you. Goodbye. |
|-----|---|

Any of responses (a)–(c), which correctly interpret the user’s move as a CR, might be regarded as useful: response (d), which incorrectly interprets it as a request or command, would not. However, without the capability to process CRs, (d) is the most likely interpretation: it is syntactically and contextually plausible (although an advanced system might be able to use intonation to help rule it out). Furthermore, correctly deciding which of (a)–(c) is most helpful depends on determining the intended *reading* of the CR (the exact question being asked).

Recent advances have been moving this way, with systems producing finer-grained CRs: (Hockey et al., 2002) clarify problematic words; (Traum, 2003) can clarify phrase reference, and in a general enough way to apply to user CRs. However, these are (so far at least) restricted to particular phenomena such as NPs, and tend to treat CRs as governed by different rules (even different processing modules) from standard dialogue (see section 6 for discussion). However, CRs are clearly not restricted to NPs, and would benefit from a general approach integrated seamlessly within the dialogue.*

This paper outlines an approach to CRs which is fine-grained enough to include specific CR questions about individual words and phrases, and general enough to handle arbitrary phrase types for both user and sys-

* See the following imagined example from (Stone, 2003b) for some non-NP CRs:

tem CRs. A prototype information-state (IS)-based system, CLARIE, is described which shows how the approach can be implemented in principle and governed by some basic empirical findings. Importantly, CRs are treated as part of the general dialogue management process – they are analysed as standard utterances, processed within the general approach to fragment resolution (albeit with a particular kind of contextual dependence), and convey standard *ask* moves, although they ask CR questions.

This is achieved via a view of utterances as *contextual abstracts* requiring a *grounding* process to fully specify their content; a radically contextualised semantic representation including a view of ellipsis as abstraction; and a simple set of pragmatic contextual operations implemented as IS update rules. The system itself is implemented using the TrindiKit (Larsson et al., 2002), building upon the GoDiS dialogue system (Larsson et al., 2000). Being intended only as a proof of concept, it is currently text-based and has only a toy lexicon and domain – although it should be extendable in future. This paper will concentrate on the novel semantic representation and the grounding process which enable its clarificational capabilities.

2. Background

CR Phenomena CRs can take many different surface *forms*, from full reprise sentences (3), via elliptical fragments (4) and sluices (5), to highly conventionalised particles (6). They can ask about a whole utterance (6) or a specific sub-constituent (5). They can also be ambiguous between various *readings*: (4) could be taken as asking either a question paraphrasable as “*Is it really (only) **two people** you’re telling me are in the class?*”, or instead “*Who are the ‘two people’ you speak of?*”. In (4), the response shows that the first interpretation was taken, but the second is equally possible in other contexts. Note that these are genuinely different questions, expecting entirely different answers (the first, a *yn*-question, can be answered “*Yes*” or “*No*”, whereas the second is effectively a *wh*-question, requiring an answer about the two people’s identities) and often given different intonations. In these examples and throughout the paper, CRs are shown **bold**, with their queried subconstituent (or *source*) shown underlined.*

- (3) | Orgady: I spoke to him on Wednesday, I phoned him.
 | Obina: **You phoned him?**
 | Orgady: Phoned him.

- (2) | Q: What do I do next? A: Slide the sleeve onto the elbow.
 | Q: What do you mean *sleeve*? A: That tube around the pipe at the joint.
 | Q: What do you mean *slide*? A: Just push the sleeve gently over along the pipe.
 | Q: What do you mean *onto*? A: The sleeve can hang there safely out of the way.

* Unless otherwise stated, all examples are from the BNC – see (Purver et al., 2003a).

- (4) | Lara: There's only two people in the class.
 | Matthew: **Two people?**
 | Unknown: For cookery, yeah.
- (5) | Sarah: Leon, Leon, sorry she's taken.
 | Leon: **Who?**
 | Sarah: Cath Long, she's spoken for.
- (6) | Anon 2: Gone to the cinema tonight or summat.
 | Kitty: **Eh?**
 | Anon 2: Gone to the cinema

While there has been extensive research into the possible levels of information which CRs can query (e.g. (Larsson, 2002; Gabsdil, 2003; Schlangen, 2004) but going back at least to (Clark, 1996)), there has been less which examines their possible forms, and the relation between this surface form and the reading (the question being asked). Two such recent studies are PGH and R&S.

PGH used a c.150,000 word portion of general dialogue from the British National Corpus (BNC) (Burnard, 2000). They identified 418 CRs and classified them by *reading* (the type of question asked) and *form* (the surface form), deriving the classifications of tables I and II, which covered about 99% of the sub-corpus.* The *conventional* form corresponds to example (6) above; the *literal reprise* to example (3); the *reprise fragment* to example (4) and the *reprise sluice* to example (5). The *non-reprise* and *wh-substituted reprise* forms should be self-explanatory, but the *reprise gap* may be less clear; it asks about the word *following* an echoed phrase, as in (7):

- (7) | Laura: Can I have some toast please?
 | Jan: **Some?**
 | Laura: Toast

As for readings, the *clausal* and *constituent* classes correspond to the two possible questions described above for example (4); the *lexical* reading concerns the identity of an uttered word, rather than its content.

R&S looked at 230 CRs in a c.35,000 word corpus of German task-oriented dialogue (SFB-360, 2000). While their classification scheme differs, the results seem to agree with PGH in most ways, including the overall frequency of CRs, and their most common forms. Both PGH (partly in follow-up work (Purver et al., 2003b; Healey et al., 2003; Purver, 2004)) and R&S then show strong correlations between some of the forms and readings. For example, the *wh*-versions (sluices and *wh*-substituted reprises) always seem to be

* PGH give a further form *gap filler* (a suggested completion for an incomplete utterance); as this does not seem relevant for computational dialogue systems (and was very rare) it is left aside here. They also treat *correction* as a further reading; this is regarded here as a separate (although closely related) phenomenon, discussed in section 6.

Table I. Possible CR Forms

<i>Form</i>	<i>Example</i>
Conventional	A: “Did Bo leave?” B: “Eh? / What? / Pardon?”
Non-Reprise	A: “Did Bo leave?” B: “What did you say? / Did you say ‘Bo’?”
Literal Reprise	A: “Did Bo leave?” B: “Did BO leave?” / “Did Bo LEAVE?”
WH-Substituted Repr.	A: “Did Bo leave?” B: “Did WHO leave?” / “Did Bo WHAT?”
Reprise Sluice	A: “Did Bo leave?” B: “Who? / What?”
Reprise Fragment	A: “Did Bo leave?” B: “Bo? / Leave?”
Reprise Gap	A: “Did Bo leave?” B: “Did Bo . . . ?”

Table II. Possible CR Readings

<i>Reading</i>	<i>Example</i>
Clausal	“Is it Bo _i you’re asking if i left? / leaving _j you’re asking if Bo did j ?”
Constituent	“Who/what do you mean by ‘Bo’ / ‘leave’?”
Lexical	“Did you say ‘Bo’ / ‘leave’?”

expressible via the clausal reading; conventional CRs cannot, instead asking either constituent or lexical questions. Some (most importantly the reprise fragment form, the most common after the conventional form) are highly ambiguous, taking any reading; although part-of-speech, surface parallelism and intonation can all help disambiguate (see below).

The two classification schemes are inspired by different underlying formalisms: R&S follow (Schlangen, 2004)’s view of CRs being caused by problems at various grounding levels (acoustic, lexical, semantic etc.); while PGH assume (Ginzburg and Cooper, 2004)’s view of CRs as being caused by individual contextually dependent parameters of utterances, and problems in their grounding. The two views seem reconcilable (see section 6); but PGH’s scheme is taken here, mainly because its parameterised approach allows CR forms and readings to be linked specifically with grammatical analysis of the CR. We now turn to this analysis.

Utterances as Contextually-Dependent Abstracts Itemizing these forms of CR only goes so far – to deal with them in a dialogue system we need to give them a suitable linguistic representation, and specify how it can be built up systematically. A suitable analysis of CRs must provide two things: a representation for normal utterances which explains how and why they can cause CRs; and a compatible representation for CRs themselves (including their often elliptical forms). (Ginzburg and Cooper, 2004, hereafter G&C) provide a HPSG analysis of CRs that promises both. Utterances are represented as functions from context to content (encoding Montague (1974)’s *meaning* or

Kaplan (1989)'s *character* rather than fully specified *content*). Contextually dependent parameters such as the reference of proper names (as well as speaker, hearer and utterance time) are abstracted to a set expressed in HPSG terms as a C-PARAMS feature. (8) shows this for an utterance “*I want to go to Paris*”:^{*}

$$(8) \left[\begin{array}{l} \text{PHON} \quad \langle \text{i, want, to, go, to, paris} \rangle \\ \text{CONTENT} \quad [\text{assert}(A, B, \text{want}(A, \text{go-to}(A, X)))] \\ \text{C-PARAMS} \quad \{ [A: \text{speaker}(A)], [B: \text{addressee}(B)], [X: \text{name}(X, \text{paris})] \} \end{array} \right]$$

Here, the value of the PHON attribute represents the surface string; CONTENT the semantic content;^{*} and C-PARAMS the abstracted set, explicitly expressing the utterance's contextual dependence. These representations are interpreted as simultaneous λ -abstracts, specifically *simultaneous abstracts with restriction* as shown in (9).

$$(9) \lambda\{ABS\}[RESTR].BODY$$

Here, $\{ABS\}$ is the set of abstracted indices, $[RESTR]$ a set of restrictions which must be satisfied during application, and $BODY$ the body of the abstract (in this case, the semantic content).^{**} For (8), the corresponding abstract would be (10), or simplifying by removing the parameters for speaker and addressee as will be done hereafter, (11). Here and throughout, a Prolog-style convention is followed, showing abstracted parameters in upper-case; their instantiated equivalents will be lower-cased:

$$(10) \lambda\{A, B, X\}[spkr(A), addr(B), name(X, paris)].assert(A, B, want(A, go-to(A, X)))$$

$$(11) \lambda\{X\}[name(X, paris)].assert(want(go-to(X)))$$

An addressee must then apply this abstract to the context, instantiating the content by finding a suitable referent for X which satisfies the restriction $name(X, paris)$. G&C term this process *grounding*.[‡] If this cannot be done

^{*} The representation $[x: restr(x)]$ is shorthand for G&C's notion of *parameter* – a referential index x paired with a restrictive fact $restr(x)$. Note that AVM (8) is simplified in several ways for clarity; in particular, syntactic information is omitted, and the use of a *go-to* predicate ignores details of the representation of verbs and modification.

^{**} Note that the content includes the illocutionary predicate *assert* and therefore expresses the conversational move (dialogue move) made by the utterance (see (Ginzburg et al., 2003) for details). This simplifies the coercion operations required in section 4.

^{**} For further formal details, see (Ginzburg and Sag, 2000).

[‡] The term *grounding* is often used in a wider sense to cover the general process of understanding and addition to the common ground. G&C use it narrowly to refer to the instantiation of abstracted parameters in context. We follow their usage here; as developed below, the parameters cover many levels of grounding including utterance form and contextual information as well as traditional sentential semantics. It is not taken to

(e.g. the hearer does not know what Paris is, or instantiating X to the known Paris makes this assertion inconsistent with previous beliefs), the utterance cannot be grounded and this can lead to a CR concerning the intended reference of the specific problematic parameter [$X : name(X, paris)$].

CR Interpretation via Contextual Coercion The resulting CR may take many forms, one of the most common being the reprise fragment “Paris?” (although as we have seen others are possible). Giving a full interpretation to these forms is a challenge: fragments and sluices require a suitable fully propositional interpretation; and even apparently fully sentential versions such as the *wh*-substituted reprise “You want to go where?” must be interpreted as asking a question about the intended previous assertion “Where is it you’re telling me you want to go?”, rather than the direct “Where do you want to go?” which may have other answers.

G&C provide an analysis using a *question-under-discussion* (QUD)-based approach to ellipsis. Intuitively, dialogue moves raise questions (QUDs) in context, and these are used to resolve subsequent fragments: asking a question “Who likes Mary?” raises it as QUD, allowing a bare answer “John” to be understood as meaning *John likes Mary*; asserting a proposition “Someone likes Mary” raises the question of *whether someone likes Mary* as QUD, licensing acknowledgements, (dis)agreements or follow-up questions “Who?”. Fragments are given an underspecified content constrained to be dependent on the currently contextually maximal QUD, encoded in their HPSG representation via a MAX-QUD feature.* Here and throughout, questions will be represented as abstracts: *wh*-questions as $?X.p$ (where X is a queried *wh*-element abstracted from a propositional body p), *yn*-questions as $?p$ (an abstract with an empty abstracted set – see (Ginzburg and Sag, 2000)). These can be thought of as λ -abstracts; in particular, the application of a question $?X.p$ to a particular (answer) value y will yield $p[X/y]$, i.e. p with all occurrences of X replaced by y . So, asking a question “Who likes Mary?” raises $?X.like(X, m)$ as the maximal QUD. The content of a fragment answer “John” is the application of the current QUD $?X.p$ to

cover e.g. the speaker’s underlying intentions, though; some readers might therefore prefer to think of it as *grammatically related grounding*.

* G&C also use a SAL-UTT feature to enforce syntactic parallelism – this is ignored here as are all syntactic matters, but can be incorporated: see (Purver, 2004).

some referent j , and this is $like(X, m)[X/j] = like(j, m)$. This is illustrated in (12) including the added complication of parameter grounding:

$$\begin{array}{ccc}
 \textit{Original question:} & & \textit{Resulting grounded context:} \\
 \left[\begin{array}{l} \text{PHON} \quad \text{who, likes, mary?} \\ \text{C-PARAMS} \quad \left\{ M : name(M, mary) \right\} \\ \text{CONTENT} \quad ask(?X.like(X, M)) \end{array} \right] & \xrightarrow[\text{[M/m]}]{\text{ground}} & \left[\text{MAX-QUD} \quad ?X.like(X, m) \right] \\
 (12) \quad \textit{Elliptical answer:} & & \downarrow \\
 \left[\begin{array}{l} \text{PHON} \quad \text{john} \\ \text{C-PARAMS} \quad \left\{ J : name(J, john) \right\} \\ \text{CONTENT} \quad assert(p[X/J]) \\ \text{MAX-QUD} \quad ?X.p \end{array} \right] & \xrightarrow[\text{[J/j]}]{\text{ground}} & \left[\text{CONTENT} \quad assert(like(j, m)) \right]
 \end{array}$$

In (12) the QUD is raised explicitly by the antecedent *ask* move. However, in the case of CRs, the relevant QUD question is not explicitly raised; instead, G&C take failure of grounding for a particular C-PARAMS parameter to license one of a set of generally available *coercion operations* (part of an interlocutor’s dialogue capability). These produce an updated context with a new (clarificational) QUD, which can resolve the CR fragment as asking a question concerning the problematic parameter. The operations are equally available to speakers and addressees of CRs; but while the chosen parameter and operation are known for speakers, addressees must presumably somehow determine which are intended (although they do not specify how). They give two specific such operations, termed *parameter focussing* and *parameter identification*, which lead to different contexts and thus different reprise readings. The first applies when grounding produces surprising or inconsistent content: the new contextual QUD becomes (roughly) the question “For which X did you say you want to go to X ?”, resolving an elliptical CR “Paris?” as asking the yes/no question “Is it really **Paris** you are saying you want to go to?”, i.e. PGH’s *clausal* reading. The second applies when no referent can be found: it produces a context where the QUD, and the resolved content of the CR, is the *wh*-question “What do you intend the word ‘Paris’ to refer to?”, i.e. the *constituent* reading. In this second case, the question refers to the previous utterance *Paris* and its intended content; utterances must therefore be included in the semantic framework, and the elliptical fragment given an *utterance-anaphoric* analysis. More detail is given below.

3. Utterance Representation

G&C’s analysis applies (explicitly at least) only to proper names. This section extends their general approach to cover a wide range of word and phrase

types and a wide range of CR forms, together with an integrated account of ellipsis and reprises.

Radical Contextual Abstraction Given the view of clarification as querying contextually dependent abstracted parameters, a suitable semantic representation must require *all* those elements of an utterance with clarificational potential (i.e. that can function as *sources* of CRs) to be included in the abstracted set. This leads to a radically contextualised representation. Firstly, to explain PGH and R&S’s data, the abstracted set must include the denotations not only of proper names, but of definite NPs, common nouns, verbs and even function words such as determiners, as any of these can be subsequently clarified (and can be seen as contextually dependent). (13) shows the effect for “*The dog snores*”:

$$(13) \lambda\{X, Q, D, S\}.[X = Q(D), Q = \textit{the}', \textit{name}(D, \textit{dog}), \textit{name}(S, \textit{snores})]. \\ \textit{assert}(S(X))$$

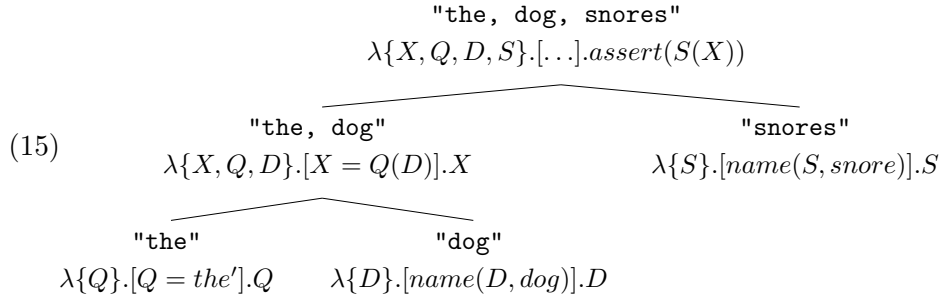
Nouns and verbs denote named predicates (D, S); determiners denote logical relations (Q). NPs are treated as lower-order, denoting sets of individuals (X) rather than generalised quantifiers (the sets being the *witness sets* of the equivalent generalised quantifiers (Barwise and Cooper, 1981)). For definite NPs, these sets are contextually abstracted; with indefinites and other quantifiers they are existentially quantified within the utterance.* Secondly, the set must also contain the subconstituents associated with each semantic parameter, as their identity must also be established and can be clarified. As we concentrate on semantic and pragmatic aspects here, utterance parameters will only be shown when necessary hereafter, but should be assumed (for more detail see (Ginzburg, forthcoming)):

$$(14) \lambda\{X, U_X, \dots\}.[X = Q(D), \textit{content}(U_X, X) \wedge \textit{phon}(U_X, \textit{"the dog"}), \dots] \dots$$

The overall representation must be built up from the input string by a suitable processing module. “Suitable” here means that it must meet the following criteria: firstly, that all potentially clarifiable parameters are contextually abstracted; secondly, that a record of sub-constituents is kept along with their contents and surface strings; and thirdly, that each of these sub-constituents be associated with all and only the contextual parameters which it contributed itself. This will allow us to ensure that a CR clarifying a particular sub-constituent can ask only about that constituent’s contributions to the utterance.

* Mass nouns and bare plurals are more complex, seen as ambiguous between predicates (or kinds) and existentially quantified sets. Quantifier scope is treated via a functional analysis; monotone decreasing quantifiers via a representation as pairs of sets; intrasentential anaphora by allowing otherwise contextually dependent items to bind to intrasentential referents during parsing – see (Purver, 2004).

This is illustrated in (15), but note that the method and formalism used are not critical as long as these requirements are met.** In the current CLARIE system, there are two alternative processing modules available: a HPSG grammar extended from (Fernández et al., 2004a) and a simpler chunk-parsing alternative. As both produce suitable contextually-dependent representations, it is not important for the purposes of the current paper which is used, and the details here will abstract away from either implementation. We will assume that for any utterance u its subconstituents s are available (written $sub-constit(u, s)$), together with their (contextually abstracted) content $content(s, c)$ and their surface string $phon(s, p)$.



Ellipsis & Reprises via Abstraction Following G&C, the content of reprises (and other elliptical fragments) is specified as dependent on a maximal QUD provided by the context. However, their approach is modified here by making this further dimension of contextual dependence explicit: the MAX-QUD feature* is also taken to be a member of the utterance’s abstracted set. A fragment “Paris” is therefore represented as in (16): its content will be an assertion of a proposition concerning some object X named Paris, but first not only X but a maximal QUD Q must be found in context to fully specify that proposition and the role of X in it:

$$(16) \lambda\{X, Q\}.[name(X, paris), max-qud(Q) \wedge (Q = ?Y.p)].assert(p[Y/X])$$

Fragments are therefore contextual abstracts, just like standard utterances; they are just *more* contextually dependent, including an extra QUD parameter.** Their resolution no longer has to be performed by a separate module: as all contextual dependence is expressed together, it becomes part of the grounding process, instantiating all parameters to obtain the fully specified content. This also avoids some potential problems (Schlangen,

** The requirements are not trivial; an HPSG approach requires several departures from standard HPSG views on semantic inheritance (Purver and Ginzburg, 2004; Purver, 2004).

* We are ignoring SAL-UTT here, but the same modification applies – see (Purver, 2004).

** Note that this use of abstraction in ellipsis differs from that of (Dalrymple et al., 1991) for VP ellipsis, in which abstracts are formed from the antecedent and used in resolving the ellipsis. Here, the elliptical fragment is the abstract, to be applied to the context.

2003) with G&C’s analysis: the fragment becomes a well-defined object (a simultaneous abstract) rather than being left underspecified (potentially problematic for standard grammars and/or parsers), and can be derived entirely compositionally, with the body derived from the constituent words and the abstracted set expressing only its contextual dependence (specifying the type of context that the abstract can be applied to).

Different fragment types correspond to different constraints on the contextual QUD. (16) above shows a standard declarative fragment (e.g. a bare answer), whose content is an assertion of a QUD-dependent proposition. (17) shows an interrogative equivalent “*Paris?*”; here the fragment asks a question (again, dependent on the QUD). In this case it is a *yn*-question; exactly the same specification can apply to *wh*-fragments such as sluices, and in those cases their lexical semantics will cause it to be a *wh*-question.

$$(17) \lambda\{X, Q\}.[name(X, paris), max-qud(Q) \wedge (Q = ?Y.p)].ask(?p[Y/X])$$

In fact, as long as CRs are taken to be standard *ask* moves, whose question happens to concern a previous utterance, this also allows us to interpret (and generate) interrogative fragments as CRs. Correct resolution (for either speaker or addressee) will depend only on producing a CR-relevant QUD in context (see the next section). And this approach need not be restricted to non-sentential fragments, but applies equally to sentential reprises (e.g. “*You want to go to Paris?*”). Sentences can be given a representation (18), again asking about some QUD-dependent proposition p , now with p constrained to be a conversational move (e.g. *ask* or *assert*) concerning the sentence’s standard propositional content:^{*}

$$(18) \lambda\{X, W, G, Q\}.[name(X, paris), name(W, want), name(G, go-to), \\ max-qud(Q) \wedge (Q = ?Y.p) \wedge (p = i(W(G(X)))) \wedge illoc-pred(i)]. \\ ask(?p[Y/Z \in \{X, W, G, Q\}])$$

This constrains the QUD to be a CR question concerning an utterance with similar propositional content, differing maximally in the QUD-abstracted parameter Y – a reprise sentence cannot ask about a semantically irrelevant antecedent. As with fragments, the lexical content of any *wh*-elements in the sentence would ensure that the overall question asked is a *wh*-question.

Finally, and again following G&C, any utterance may be seen as utterance-anaphoric. Here, the word *Paris* is not taken to denote an object named

* In showing abstraction of some Y (possibly a predicate) from $i(W(G(X)))$, (18) may suggest that full higher-order abstraction is available. This is not the case – the expression is encoded as a Prolog term from which only specific elements can be abstracted.

Paris as before, but a previous utterance U of ‘*Paris*’; the QUD is therefore constrained to be a question about such an utterance, as in (19):

$$(19) \lambda\{Q\}.[max-qud(Q) \wedge (Q = ?Y.P(U, Y)) \wedge phon(U, "paris")].ask(Q)$$

Of course, this approach means a threefold ambiguity in the representation given to most utterances: for full sentences, a standard and a reprise version; for fragments, a declarative and an interrogative version; and for both of them an utterance-anaphoric version. The declarative/interrogative distinction might be removed by treating the illocutionary predicate as a member of the contextually-abstracted set (see (Ginzburg, forthcoming) for a proposal), but the utterance-anaphoric distinction seems fundamental and necessary. Note that the conditions on the abstracted parameters will determine which of the possible representations can apply to a given context, allowing context to disambiguate in many cases; but where ambiguity persists, we need other disambiguation methods. This is the province of the grounding mechanism, described in the next section.

The generality of the approach is important: utterances are seen as contextually dependent abstracts, with their abstracted parameters defining their clarificational potential; fragments as utterances with slightly more contextual dependence; and CRs as fragments with specific kinds of contextual dependence. They are therefore treated as standard interrogative *ask* moves, with their CR nature coming only from the question raised in their context. They also exhibit standard contextual dependence, of course, with lexically contributed parameters requiring grounding; CRs-of-CRs are therefore possible, and do occur in corpora (Purver, 2004).

4. Utterance Processing & Grounding

Handling clarificational dialogue now centres around the grounding process: application of the utterance abstract to the context to derive its fully specified content, finding suitable referents for each of the abstracted parameters (including the *max-qud* parameter if it exists). It is the inability to ground user parameters in context (or to ground them in a way that is consistent with what is already known) that gives rise to system CRs; it is the grounding of parameters in a suitable way that allows user CRs (and utterances in general) to be interpreted correctly.

It is therefore the grounding process which disambiguates between possible utterance interpretations. Much ambiguity is now represented directly via contextual parameter abstraction – lexical ambiguity, ambiguity of reference, and propositional underspecification of fragments – and this must all be resolved during grounding. However, there are other forms of ambiguity that are not represented in this way – particularly syntactic ambiguity such as propositional phrase attachment. While it may seem tempting, and may

be possible, to represent this via parameter abstraction too (e.g. by regarding PPs as dependent on abstracted event parameters), such a move does not appear to fit with observed data. Remember that abstracted parameters represent elements with *clarificational potential*, that cause CRs when their grounding fails; but no CRs querying syntactic structure were found by either PGH or R&S, and examples are hard to imagine. Syntactic ambiguity must therefore be represented in other ways. In CLARIE with its small grammar, multiple hypotheses are simply produced as a set; with larger grammars ambiguity increases, but this is a standard problem and standard approaches could be taken (efficient chart representations, ambiguity reduction via probabilistic parsing etc.).* The extra ambiguity introduced in section 3 (standard sentences vs. reprise or utterance-anaphoric versions) will add to this, but only by a known and constant factor.

Whatever the representation, we take it to be the grounding process that makes the final choice between interpretations: the choice of contextual parameter instantiation is not in principle independent from the choice between syntactic representations – particular instantiations may only make sense with particular representations, and vice versa.* Grounding must therefore both find an assignment for the abstracted parameter set *and* choose a preferred interpretation. In CLARIE, this is implemented in the simplest way possible, avoiding heavyweight reasoning or inference. Possible grounding strategies are defined as TrindiKit IS update rules tried in order, applying as soon as a possible interpretation and an assignment for its abstracted set are found that satisfy the pre-conditions. The rule pre-conditions express general constraints such as internal and contextual consistency; the rule ordering ensures maximally relevant grounding (e.g. instantiation as an answer to a question currently under discussion if possible, and only as a CR if not),** and for CRs is driven by PGH’s empirical correlations.

Utterance Processing The CLARIE information state is shown in (20), with some simplification.† The AGENDA field holds the system’s immediate intended actions, COM is a set of public commitments (represented as propositions) and BG a set of corresponding descriptions of referents (represented as parameters $[x : restr(x)]$ just as in the contextually abstracted sets of

* Another possibility for ambiguity reduction is of course incremental interpretation; combining this with the parameter grounding process seems an interesting challenge.

* Disambiguating here (rather than inside the grammar) also has the advantage of access to contextual IS information: not only possible parameter referents, but the current dialogue state (QUDs, previous moves, established beliefs etc.).

** This approach therefore assumes that contextually coherent non-CR interpretations should always be preferred. There may be cases where this does not hold in practice, such as when intonation strongly indicates a CR; see section 6.

† As before, we ignore syntactic information; also any explicit division between public and private knowledge. With the exception of AGENDA, the IS here can be taken as representing the common ground.

utterances). QUD is a stack of QUDs, used for ellipsis resolution and the determination of dialogue move (e.g. recognition of answerhood).

$$(20) \left[\begin{array}{ll} \text{AGENDA} & \langle \text{stack}(\text{action}) \rangle \\ \text{COM} & \langle \text{set}(\text{proposition}) \rangle \\ \text{BG} & \langle \text{set}(\text{parameter}) \rangle \\ \text{QUD} & \langle \text{stack}(\text{question}) \rangle \end{array} \quad \begin{array}{ll} \text{UTT} & \langle \text{nstackset}(4, \text{sign}) \rangle \\ \text{MOVE} & \langle \text{stackset}(\text{proposition}) \rangle \\ \text{PENDING} & \langle \text{stack}(\text{set}(\text{sign})) \rangle \end{array} \right]$$

UTT is an utterance record in linear dialogue order. This provides a record in which to find the sources of CRs, allowing user CRs to be properly interpreted and their answers to be determined. Its members must therefore be *signs*, rather than semantic representations such as *moves*: they must include all information which may be clarified by the various possible CR readings (including lexical identity), and maintain sub-constituent structure as outlined in section 3. These are large and complex representations; to simplify matters and reduce computational overhead, UTT can have a limited length, 4 utterances in the current implementation, as PGH found that CRs beyond this distance are rare (this could of course be extended as desired). A longer (unlimited) record of the purely semantic content of the dialogue is therefore kept in MOVE (the fully grounded and instantiated moves made).^{‡‡} This reflects the psycholinguistically observed difference in memory length between high- and low-level information (Fletcher, 1994). PENDING holds ungrounded utterance abstracts (in fact, ambiguous *sets* of utterance abstracts) during the grounding process.

Utterance processing is based broadly on G&C’s original protocol. First, an incoming ungrounded user utterance (set of abstracts) *U* is pushed onto the PENDING and UTT stacks, while grounding is attempted:

$$(21) \left[\begin{array}{ll} \text{AGENDA} & \langle \dots \rangle \\ \text{QUD} & \langle \dots \rangle \\ \text{MOVE} & \langle \dots \rangle \end{array} \quad \begin{array}{ll} \text{UTT} & \langle \mathbf{U} = \lambda\{X, Y\}.p, \dots \rangle \\ \text{PENDING} & \langle \mathbf{U} = \lambda\{X, Y\}.p \rangle \end{array} \right]$$

Next, the following schema is applied:

1. Try to *integrate* as a standard utterance in the current context.
2. Try to *accommodate* into the context using domain-specific knowledge.
3. Try to interpret as a CR via a *coercion* operation on context.
4. Fail, and ask a relevant CR for *clarification*.

^{‡‡} The ordering of MOVE will not necessarily reflect linear dialogue order, but the order of grounding and therefore the order in which moves take effect.

The standard *integration* rules use the current IS (the members of BG and the maximal member of QUD) to provide a parameter assignment. If integration fails, *accommodation* rules try to ground using a new plausible, but not explicitly asked, QUD determined from the domain plan: in a travel-agent domain, allowing an utterance “*Paris, please*” to be interpreted as answering an unasked but mutually understood question “*Where are you travelling to?*”. We will leave accommodation aside here (see (Larsson et al., 2000) for details) and concentrate instead on the *coercion* rules, which attempt to ground the utterance as a user CR. Again, a new relevant QUD is produced to ground the abstracted *max-qud* parameter, but rather than domain knowledge, a general set of contextual coercion operations are used which produce CR-related QUD questions. In all cases, success means that the grounded (fully specified) utterance U is removed from PENDING and its update effects applied to the IS, minimally adding the move that it makes to MOVE, but also including other move-specific effects. If coercion rules have grounded U as a CR, its resolved content p is a move $ask(q)$, and its effects raise q as the new QUD, and an action to respond. U remains in the linear utterance record UTT, of course – future CRs may ask about it:

$$(22) \left[\begin{array}{ll} \text{AGENDA} & \langle \text{respond}(q), \dots \rangle \\ \text{QUD} & \langle q, \dots \rangle \\ \text{MOVE} & \langle p[X/x, Y/y] = ask(q), \dots \rangle \end{array} \quad \begin{array}{ll} \text{UTT} & \langle U = \lambda\{X, Y\}.p, \dots \rangle \\ \text{PENDING} & \langle \rangle \end{array} \right]$$

If all grounding rules fail, the utterance cannot be grounded in the current IS, and a set of *clarification* rules use its ungrounded parameters to form a suitable CR (about one particular problematic parameter, or about the entire utterance) and add an agenda action to ask it. Importantly, the ungrounded utterance is left in PENDING so that grounding can be re-attempted if a subsequent user answer to the CR provides sufficient new information:

$$(23) \left[\begin{array}{ll} \text{AGENDA} & \langle \text{clarify}(U, X), \dots \rangle \\ \text{QUD} & \langle \dots \rangle \\ \text{MOVE} & \langle \dots \rangle \end{array} \quad \begin{array}{ll} \text{UTT} & \langle U = \lambda\{X, Y\}.p, \dots \rangle \\ \text{PENDING} & \langle U = \lambda\{X, Y\}.p \rangle \end{array} \right]$$

Grounding Standard Utterances Standard *integration*, then, must instantiate parameters in the current IS unchanged. The *max-qud* parameter must be identified with the head of the QUD stack; parameters concerning individual referents (of names and definites) with unique antecedents in the BG set or available from the domain model; and parameters concerning

predicates (from nouns and verbs) from BG or the lexicon.* Logical relations (contributed by closed-class function words) are always available from the lexicon. A simple consistency check is then applied, preventing grounding of e.g. assertions that conflict with previously held beliefs (which therefore lead to clarification – see below). The current implementation checks constraints separately: unique referents must be found for each parameter in turn, with dependent parameters ordered where possible (if *a* depends on *b*, *b* must be found first). This is sufficient in the current simple domain; more complex domains will require simultaneous satisfaction (Mellish, 1985; Haddock, 1987). The protocol is as follows:

1. Interpret as answering the currently maximal QUD question.
2. Interpret as asking a question which is relevant to the current IS.
3. Interpret as a greeting, closing or thanking move.

Success, with all abstracted parameters being found unique referents which fit their required constraints, removes the utterance from PENDING and applies its update effects. These follow GoDiS in general, with *ask* moves adding a new question to QUD, and a new action to answer it to AGENDA; answers (*assert* moves which answer a question in QUD) remove the answered question and add a new belief to COM.

Grounding User CRs In contrast, *coercion* modifies the context, producing a new QUD to ground the *max-qud* parameter. There are four defined coercion operations (two directly from G&C, and two new similar versions), all with the same general form.* Given a *source* utterance (the utterance being asked about), they license a new QUD concerning a sub-constituent related to one of its contextually abstracted parameters. As addressee of a user CR, the system does not of course know the source utterance and sub-constituent or the correct coercion operation, but must identify them: an (ordered) constraint satisfaction approach is taken, searching through the UTT record most recent first (both PGH and R&S found that the vast majority of CRs queried sources from the immediately preceding turn, with numbers decreasing with distance) until a suitable source is found. This must satisfy the constraints imposed by the semantics of the CR utterance itself, the coercion operation being tested, and some general rule pre-conditions, with these pre-conditions and the rule ordering derived from empirical data. The first two coercion operations are taken from G&C's analysis of the

* This allows a degree of *alignment* behaviour (Pickering and Garrod, 2004) – previously grounded words and phrases will have referents in BG which can be used to ground subsequent uses before lexicon lookup, leading to the same interpretation where consistent.

* In (Purver, 2004), a fifth coercion operation is given for certain non-reprise and conventional CRs which have their propositional content specified directly by the grammar. The four operations given here subsume those cases, with the grammatically specified content acting as a constraint on which operation can apply.

clausal and constituent readings respectively; they term them *parameter focussing* and *parameter identification*.

Parameter Focussing This operation forms a QUD by abstracting one parameter from the overall source utterance’s content (including its conversational move predicate):

$$(24) \quad u = \lambda \left\{ \dots, X, \dots \right\} . [\dots] . p \quad \text{where} \quad p = i(p') \wedge \text{illoc-pred}(i) \\ \Rightarrow \quad \text{max-qud}(q') \wedge (q' = ?X.p)$$

Given a source utterance and a reprise CR with consistent contents, this can produce the clausal reading. Assuming a source utterance “*I want to go to Paris*”, its (simplified) representation will be as in (25):

$$(25) \quad \lambda \left\{ X, W, G \right\} [\text{name}(X, \text{paris}), \text{name}(W, \text{want}), \text{name}(G, \text{go-to})] . \\ \text{assert}(W(G(X)))$$

One possible application of the coercion operation in (24) to this source utterance is (26), abstracting the parameter associated with the word *Paris*:

$$(26) \quad \lambda \left\{ \dots, X, \dots \right\} . [\dots, \text{name}(X, \text{paris}), \dots] . \text{assert}(W(G(X))) \\ \Rightarrow \quad \text{max-qud}(q') \wedge (q' = ?X.\text{assert}(W(G(X))))$$

This new QUD q' is a question paraphrasable “*Which X is it you are saying you want to go to?*” Now, remember that one possible representation of a new incoming fragment “*Paris?*” (a potential but as yet ungrounded CR) is (17), repeated here as (27). And as shown, its *max-qud* parameter can be grounded to q' to give the desired CR interpretation, the question “*Are you telling me you want to go to Paris?*”:

$$(27) \quad \lambda \left\{ X, Q \right\} . [\text{name}(X, \text{paris}), \text{max-qud}(Q) \wedge (Q = ?Y.p)] . \text{ask}(?.p[Y/X]) \\ \text{grounding } [Q/q'] \quad \Rightarrow \quad \text{ask}(?.\text{assert}(W(G(X))))$$

Note that other applications of (24)’s coercion operation to the source utterance (25) are possible (e.g. abstracting the predicate parameters associated with *want* or *go to*), but these will produce QUDs which are not applicable to the fragment (27) as their semantic type will be inconsistent (we assume a semantic ontology which distinguishes entities from predicates and relations, as well as broad sub-classes such as human/non-human entities, definite/indefinite determiner relations etc.). Similarly, other representations of the CR fragment are possible, but they will not be applicable to this QUD. Note also that there is no requirement for the CR fragment to have the same surface form or semantic referent as the original source sub-constituent *Paris*, as long as its semantic type is consistent. So non-identical CRs “*The capital of France?*” and ‘incorrect’ CRs “*Prague?*” can be analysed the same way – see example U8 in section 5 for more detail. Finally, note that the same mechanism and QUD q' can apply to CRs other than fragments (e.g. the full sentential reprise of (18) and parallel *wh*-versions).

Parameter Identification The second coercion operation applies in a similar way to yield the constituent CR reading; the new QUD is a question about the intended content of the source sub-constituent u :

$$(28) \quad u = \lambda \left\{ \dots, X, \dots \right\} . [\dots, \text{content}(u, X), \dots] . X \\ \Rightarrow \quad \text{max-qud}(q') \wedge (q' = ?X.\text{content}(u, X))$$

Applying this to our example, the source sub-constituent *Paris*, will give:

$$(29) \quad u = \lambda \left\{ X \right\} . [\text{name}(X, \text{paris})] . X \quad (\text{where } \text{phon}(u, \text{"paris"})) \\ \Rightarrow \quad \text{max-qud}(q') \wedge (q' = ?X.\text{content}(u, X))$$

Recalling the utterance-anaphoric fragment representation of (19), repeated here as (30), we can now see that this QUD is applicable (being a question about an utterance whose surface form is ‘*Paris*’), and yields an overall CR question paraphrasable “*What is the content of your utterance ‘Paris’?*”.

$$(30) \quad \lambda \left\{ Q \right\} . [\text{max-qud}(Q) \wedge (Q = ?Y.P(U, Y)) \wedge \text{phon}(U, \text{"paris"})] . \text{ask}(Q) \\ \text{grounding } [Q/q'] \quad \Rightarrow \quad \text{ask}(?X.\text{content}(u, X))$$

In this case, phonological parallelism *is* enforced by the constraints on the fragment in (30); the constituent reading is only available for fragments which echo the source exactly. This coercion operation can also apply to some non-fragment CR types, principally the conventional “*What*” type; the representation given to these by the grammar does not of course enforce such parallelism (although they have other constraints, for example to ensure that they ask about entire utterances rather than sub-constituents).

Lexical Readings Completing coverage of PGH’s reading classification requires two further coercion operations; both produce QUDs concerning lexical form. The first, *lexical identification*, queries the identity of an echoed word (licensing lexical CRs); the second, *gap identification*, queries the identity of the word following the echoed word (licensing reprise gap CRs). Both take a similar form to parameter identification above, but as both concern utterance parameters, they produce questions about utterance identity (surface form) rather than content:

$$(31) \quad u = \lambda \left\{ \dots, U_X, \dots \right\} . [\dots, \text{sub-constit}(u, U_X), \dots] . p \\ \Rightarrow \quad \text{max-qud}(q') \wedge (q' = ?U_X.\text{utter}(U_X))$$

$$(32) \quad u = \lambda \left\{ \dots, U_X, U_Y, \dots \right\} . [\dots, \text{consec-sub-constits}(u, \langle U_X; U_Y \rangle), \dots] . p \\ \Rightarrow \quad \text{max-qud}(q') \wedge (q' = ?U_Y.\text{utter-consec}(U_X; U_Y))$$

In both cases, these QUDs will apply to the utterance-anaphoric analysis of a CR fragment (30), provided that the phonological parallelism constraint is again met. Thus a fragment “*Paris?*” can be resolved as having a lexical

CR reading, asking “*Did you say the word ‘Paris’?*”, or a lexical gap reading “*What did you say **after** the word ‘Paris’?*”.

Coercion Rule Application Both CR utterances and coercion operations impose constraints on their possible sources, but multiple solutions may still be possible. Fortunately, there are several empirical correlations between properties of the source and CR and the available readings. Re-examination of PGH’s corpus and experimental data (Purver, 2004) provides various rules to help determine which operations should apply to which source and CR cases. Firstly, constituent CRs only seem to query content phrases, not function words – this is unsurprising, as we expect the meaning of function words to be mutual knowledge. Secondly, CRs in general are rare for function words, but number determiners are an exception, often queried with clausal readings (although not constituent readings). This also implies that fragments which echo other function words should be interpreted as reprise gaps (which query the following word and not that function word itself) – and indeed, experiments confirm this. Thirdly, constituent readings are more likely on the first mention of the source word or phrase; again, we might expect questions about content to be rare once it has already been used and established in the common ground. Putting these observations together, the final CR coercion rule protocol is as follows:*

1. Perform *parameter identification* and interpret as a constituent CR *if* the source is the first mention of a content phrase fragment.
2. Perform *parameter focussing* and interpret as a clausal CR *if* the source is a content phrase or number determiner.
3. Perform *gap identification* and interpret as a lexical reprise gap *if* the source is a function word.
4. Perform *lexical identification* and interpret as a lexical CR.

As a CR is an *ask* move, it adds a new question to QUD, and an agenda action to answer it. The answer can now be established from the relevant features of the source utterance in the UTT record (already identified as part of the coercion rule application): for clausal and constituent CRs, from its semantic content; for lexical or gap CRs, directly from its identity. Answers can be conveyed by standard system *assert* moves, realized via the grammar in the same way as answers to normal questions.**

* Being data-driven, these correlations may be domain- or language-dependent (clarifying some function words might be important in particular tasks). They are thus not specified within the grammar or coercion operations themselves, but as constraints on the (system-dependent) grounding rules. The current system also only uses a few of the clearest correlations.

** Generation is currently simplistic, giving no regard to user or context, with two special cases: CRs about word content are answered with alternative descriptions wherever possible rather than the original problematic form; when answering *yn*-questions negatively, an over-answer is produced by answering the new coerced QUD as well as the explicit

Grounding Failure and System CRs If a user utterance cannot be grounded in any way, the *clarification* rules produce a system CR (using the same coercion operations). Again, rules are derived from PGH’s empirical data (Purver, 2004), so that a specific CR can be phrased so that it is most likely to be interpreted correctly by the user. Particular parameter grounding problems lead to particular questions being asked: unresolved predicate parameters (i.e. out-of-vocabulary nouns and verbs) lead to constituent *wh*-questions; parameters whose instantiation causes inconsistency lead to clausal “check” *yn*-questions. Particular source subconstituent types will also lead to particular CR forms: some phrase types are most often clarified using sluices, some with reprise fragments. The overall protocol is as follows:

1. Parse failure (nothing to ground): constituent CR of whole utterance.
2. Unknown parameter (no unique referent): clausal *wh*-question about source constituent *if* a definite or pronoun (realised as a sluice); constituent *wh*-question otherwise (realised as a fragment).
3. Inconsistent parameter (grounding causes inconsistency): clausal *yn*-question, realised as a clausal reprise fragment.
4. Inconsistent moves (only grounded inconsistently) and irrelevant moves (no *max-qud* referent): constituent question about whole utterance.

As system CRs (again, being *ask* moves) introduce a new question to QUD, subsequent user answers (elliptical or not) can be interpreted according to standard answerhood rules – no special treatment is required – and as long as such answers provide the required information, the problematic PENDING utterance will now be groundable without requiring repetition.

5. Examples

The dialogue below shows some of CLARIE’s current capabilities:

S1> Where do you want to go to?		
U1> pardon?	-->	S6> How do you want to travel?
S2> Where do you want to go to?		U6> by pullman
U2> i want to go to that city		S7> 'Pullman'?
S3> Which city?		U7> train
U3> paris		S8> The trip costs fifty pounds.
S4> How do you want to travel?		U8> sixty?
U4> no i want to go to london		S9> No, fifty.
S5> London?		U9> fifty?
U5> yes london -----		S10> Yes.

Note that both system CRs (S3, S5, S7) and user CRs (U1, U8, U9) are included, as are questions about utterance form (U2), NP reference (S3) and

CR question – see example S9 in section 5. But as these are standard *ask/assert* moves, there should be no barrier to applying more advanced generation techniques when answers concern reference in complex or ambiguous situations.

noun predicate reference (S7), and both clausal (S5, U8, U9) and constituent (S7) reprises.* This section illustrates the overall processes by stepping through two examples, the user CR U8 and the system CR S7.**

Processing a User CR After generating and processing S8, the IS will appear as in (33), with this source utterance at the top of the UTT record, and its relevant (grounded) parameters in BG. Its move is an assertion of a proposition p_1 , and the question $?p_1$ has therefore become maximal in QUD.

$$(33) \left[\begin{array}{l} \text{BG} \quad \{x_1:the(trip)\} \\ \text{UTT} \quad \left\langle \left\langle \begin{array}{l} \text{S8} \left[\begin{array}{l} \text{"the, trip, costs, fifty, pounds"} \\ \lambda\{cost, 50, pound, x_1:the(trip)\}. \\ \text{assert}(sys, p_1: \exists\{x_2:50(pound)\}.cost(x_1, x_2)) \end{array} \right] \right. \\ \left. \dots \right. \end{array} \right\rangle \right\rangle \\ \text{PENDING} \quad \langle \rangle \quad \text{QUD} \quad \langle ?p_1, \dots \rangle \\ \text{MOVE} \quad \langle \text{assert}(sys, p_1), \dots \rangle \end{array} \right]$$

The incoming user utterance U8 is added to PENDING and UTT for grounding. Being a fragment, it is initially ambiguous (standard vs. utterance-anaphoric), but given the complexity of these IS representations, only the eventually successful representation is shown here:

$$(34) \left[\begin{array}{l} \text{BG} \quad \{x_1:the(trip)\} \\ \text{UTT} \quad \left\langle \left\langle \begin{array}{l} \text{U8} \left[\begin{array}{l} \text{"sixty"} \\ \lambda\{X, Q\}.[X = 60, max-qud(Q) \wedge (Q = ?Y.P)].ask(usr, ?.P[Y/X]) \end{array} \right] \right. \\ \left. \dots \right. \end{array} \right\rangle \right\rangle \\ \left\langle \left\langle \begin{array}{l} \text{S8} \left[\begin{array}{l} \text{"the, trip, costs, fifty, pounds"} \\ \lambda\{cost, 50, pound, x_1:the(trip)\}. \\ \text{assert}(sys, p_1: \exists\{x_2:50(pound)\}.cost(x_1, x_2)) \end{array} \right] \right. \\ \left. \dots \right. \end{array} \right\rangle \right\rangle \\ \text{PENDING} \quad \langle \text{U8} \rangle \quad \text{QUD} \quad \langle ?p_1, \dots \rangle \\ \text{MOVE} \quad \langle \text{assert}(sys, p_1), \dots \rangle \end{array} \right]$$

The standard integration rules fail; although the parameter X corresponding to *sixty* is groundable (available from the lexicon), the current

* The current behaviour is designed to demonstrate the elliptical CR interpretation and generation capabilities. Many system CRs (e.g. S7) may benefit from less elliptical form in practice; this behaviour can be controlled by a user-settable flag.

** IS illustrations here show grounded utterances in UTT still as abstracts, but with grounded parameters in lower-case; this can be thought of as shorthand for storing abstracts with their assignments; in CLARIE it is achieved by Prolog unification.

maximal member of QUD cannot be unified with the *max-qud* parameter Q . The coercion rules are therefore attempted; the first one, parameter identification, fails as the only possible (phonologically parallel) source is not a content phrase. The second rule, parameter focussing, *can* apply, as number determiners are acceptable; it must form a new QUD by abstracting the logical relation 50 from S8's original content, to give:

$$q = ?X.assert(sys, \exists\{x_2 : X(pound)\}.cost(x_1, x_2))$$

This can only apply to the standard (not the utterance-anaphoric) interpretation of U8 shown. Grounding the parameter Q to this new q , and X to the relation 60 as it must be, gives U8 the correct clausal CR interpretation “Is it **60** _{X} that you're telling me it costs X pounds?”:

$$(35) \left[\begin{array}{l} \text{BG} \quad \left\{ x_1 : the(trip) \right\} \\ \text{UTT} \quad \left\langle \left\{ \begin{array}{l} \text{U8} \left[\begin{array}{l} \text{"sixty"} \\ \lambda\{60, Q'\}.ask(usr, ?.assert(sys, p_2 : \exists\{x_2 : 60(pound)\}.cost(x_1, x_2))) \end{array} \right] \\ \text{S8} \left[\text{"the, trip, costs, 50, pounds"} \right], \dots \end{array} \right\rangle \right\rangle \\ \text{PENDING} \quad \langle \rangle \quad \text{QUD} \quad \langle ?.assert(sys, p_2), \quad ?.p_1, \dots \rangle \\ \text{MOVE} \quad \langle ask(usr, ?.assert(sys, p_2)), \quad assert(sys, p_1), \dots \rangle \\ \text{AGENDA} \quad \langle \text{respond}(?.assert(sys, p_2)), \dots \rangle \end{array} \right]$$

Note that the correct choice of the logical relation 50 as the parameter to abstract is required by the semantic typing assumed: Q can only be grounded to a question where the abstracted parameter is of the same type as 60 i.e. a number determiner. In more complex domains with more complex sentences, additional methods of source parameter identification may be required, such as ontological distance measures (and perhaps phonological similarity in some cases).

The new CR *ask* move is introduced to MOVE, its question to QUD, and a **respond** action to AGENDA. The answer can be determined directly from UTT (it is negative, as the question does not match the content of S8). As the question is maximal in QUD, an elliptical answer “*No*” is licensed;* once this answer has been generated, its update effects remove the answered CR question from QUD, and the dialogue continues.

Processing a System CR Here, the source utterance is U6 “*By pullman*”, and specifically the sub-constituent *pullman* which is not in the system's lexicon. (36) shows the IS after this utterance has been parsed: it is in the

* To make a negative response more natural and helpful (Hockey et al., 1997), an over-answer “*No, fifty*” is generated by forcing the coerced QUD q to be answered too.

UTT record, but also in PENDING as it has not yet been grounded. The current maximal question in QUD is the system-posed question of how the user wants to travel:

$$(36) \left[\begin{array}{l} \text{UTT} \quad \left\langle \left\{ \begin{array}{l} \text{"by, pullman"} \\ \text{U6} \left[\lambda\{N, Q\} \cdot [\text{name}(N, \text{pullman}), \right. \right. \\ \left. \left. \text{max-qud}(Q) \wedge (Q = ?X_2.P)] \cdot \text{assert}(\text{usr}, P_{X_2 \mapsto N}) \right] \right], \dots \right\} \right\rangle \\ \text{PENDING} \quad \langle \text{U6} \rangle \quad \text{QUD} \quad \langle ?X_1.\text{travel-by}(X_1), \dots \rangle \\ \text{MOVE} \quad \langle \text{ask}(\text{sys}, ?X_1.\text{travel-by}(X_1)), \dots \rangle \end{array} \right]$$

As no referent for the parameter N can be found (and coercion/accommodation rules cannot help with this), the utterance is left in PENDING, and the system adds an action to the AGENDA to clarify this particular parameter and its associated sub-constituent S . This **clarify** action causes the CR “Pullman?” to be asked (S7),* discharging the action and adding its move and question to MOVE and QUD:

$$(37) \left[\begin{array}{l} \text{UTT} \quad \left\langle \left\{ \begin{array}{l} \text{"pullman"} \\ \text{S7} \left[\lambda\{S\} \cdot \text{ask}(\text{sys}, ?N.\text{content}(S, N)) \right], \\ \text{U6} \left[\text{sub-constit}(\text{U6}, S) \wedge S = \left[\begin{array}{l} \text{"pullman"} \\ \lambda\{N\} \cdot [\text{name}(N, \text{pullman})].N \right] \right], \dots \end{array} \right] \right\} \right\rangle \\ \text{PENDING} \quad \langle \text{U6} \rangle \quad \text{QUD} \quad \langle ?N.\text{content}(S, N), ?X_1.\text{travel-by}(X_1), \dots \rangle \\ \text{MOVE} \quad \langle \text{ask}(\text{sys}, ?N.\text{content}(S, N)), \text{ask}(\text{sys}, ?X_1.\text{travel-by}(X_1)), \dots \rangle \end{array} \right]$$

* The CR produced is a constituent reprise fragment, preferred when the parameter is associated with a common noun (as here, via the construction *by N*).

Now U7, “*Train.*”, is added to UTT and PENDING. It is given the usual fragment representation, requiring a *max-qud* parameter Q'' to be grounded:

$$(38) \left[\begin{array}{l} \text{UTT} \left\langle \left\{ \begin{array}{l} \text{U7} \left[\lambda\{T, Q''\}. \left[\begin{array}{l} \text{"train"} \\ \text{name}(T, \text{train}), \\ \text{max-qud}(Q'') \wedge (Q'' = ?X_3.P'') \end{array} \right]. \text{assert}(\text{usr}, P_{X_3 \mapsto T}'') \end{array} \right], \right. \\ \left. \text{S7} \left[\lambda\{S\}. \text{ask}(\text{sys}, ?N.\text{content}(S, N)) \right], \right. \\ \left. \text{U6} \left[\text{sub-constit}(\text{U6}, S) \wedge S = \left[\begin{array}{l} \text{"pullman"} \\ \lambda\{N\}. \left[\text{name}(N, \text{pullman}) \right]. N \end{array} \right], \dots \right] \end{array} \right\rangle \\ \text{PENDING} \langle \text{U7}, \text{U6} \rangle \quad \text{QUD} \langle ?N.\text{content}(S, N), ?X_1.\text{travel-by}(X_1), \dots \rangle \\ \text{MOVE} \langle \text{ask}(\text{sys}, ?N.\text{content}(S, N)), \text{ask}(\text{sys}, ?X_1.\text{travel-by}(X_1)), \dots \rangle \end{array} \right]$$

This time, as the predicate *train* is in the system lexicon, the parameter T can be grounded; so can the *max-qud* parameter Q'' as it can be unified with the current maximal QUD, a question about the intended content of S . U7 therefore gets the required answer interpretation “*By ‘pullman’, I meant ‘train.’*” and is removed from PENDING. As it answers the current maximal member of QUD, this is removed; and a belief about the content of the problematic sub-constituent S is added to COM. In turn, this new belief in COM licenses a fact [*train*:*name(train, pullman)*] to be added to BG:

$$(39) \left[\begin{array}{l} \text{UTT} \left\langle \left\{ \begin{array}{l} \text{U7} \left[\lambda\{\text{train}, Q''\}. [Q'' = ?N.\text{content}(S, N)]. \text{assert}(\text{usr}, \text{content}(S, \text{train})) \right], \\ \text{S7} \left[\lambda\{S\}. \text{ask}(\text{sys}, ?N.\text{content}(S, N)) \right], \\ \text{U6} \left[\begin{array}{l} \lambda\{N, Q\}. \left[\begin{array}{l} \text{name}(N, \text{pullman}), \\ \text{max-qud}(Q) \wedge (Q = ?X_2.P) \end{array} \right]. \text{assert}(\text{usr}, P_{X_2 \mapsto N}) \\ \text{sub-constit}(\text{U6}, S) \wedge S = \left[\begin{array}{l} \text{"pullman"} \\ \lambda\{N\}. \left[\text{name}(N, \text{pullman}) \right]. N \end{array} \right], \dots \end{array} \right] \end{array} \right\rangle \\ \text{PENDING} \langle \text{U6} \rangle \quad \text{QUD} \langle ?X_1.\text{travel-by}(X_1), \dots \rangle \\ \text{MOVE} \langle \text{assert}(\text{usr}, \text{content}(S, \text{train})), \\ \text{ask}(\text{sys}, ?N.\text{content}(S, N)), \text{ask}(\text{sys}, ?X_1.\text{travel-by}(X_1)), \dots \rangle \\ \text{BG} \left\{ [\text{train}:\text{name}(\text{train}, \text{pullman})], \dots \right\} \end{array} \right]$$

The original problematic utterance U6 is still in PENDING, but there is now enough information in the IS to ground both the parameter N associated with *pullman* (to *train*), and the *max-qud* parameter Q (to the once-again

maximal QUD asking how the user wants to travel). It is therefore given the correct interpretation (that the user want to travel by train), and the dialogue continues without this answer having to be separately given. And with the information about the alternative name of the *train* concept now in BG, subsequent uses of the word *pullman* can be interpreted without clarification – the system has effectively learnt the word. While this example’s direct user answer may seem like a simplification, note that its interpretation was achieved by standard processes (elliptical fragment resolution by QUD), rather than any CR-specific mechanism; although more complex or indirect answers will no doubt happen, this is equally true for non-CR questions, and the resolution mechanisms will be the same. This is, of course, a direct consequence of treating CRs as standard *ask* moves which ask standard questions (which happen to be about other utterances).

6. Discussion

Other Approaches Most previous approaches to CRs treat them as separate from general dialogue processes, interpretation and generation, and thus restricted to specific phenomena and system-generated CRs. (Hockey et al., 2002) highlight problematic words and suggest reformulation, but use a separate module external to the dialogue manager; (Dusan and Flanagan, 2002) clarify words and acquire their meanings, but require users to answer using specific formulations. (Kievit et al., 2001)’s approach is more similar, generating CRs from problematic context-dependent referents, and using a similar PENDING stack – but is restricted to NP reference, and resolves CR answers by an apparently CR-specific method of unifying with the source. Perhaps the most similar is (Traum, 2003); the approach fits within the dialogue processes and is general enough to apply to user CRs. Again though, it is restricted to certain forms of NP reference question and general “*Say again?*” requests (although extension would be possible); and sees CRs and their answers as dedicated *request-repair* and *repair* moves (rather than *ask/assert*), complicating any potential account of indirect answers or clarification of CRs themselves.

Other theoretical approaches have associated CRs either with parameters in the user’s plan (Litman and Allen, 1987; Heeman and Hirst, 1995) or with problems at various general levels of grounding (Larsson, 2002; Schlangen, 2004, R&S). This approach differs from both in its explicit association of problematic parameters with their source words or phrases via their grammatical representation. This covers CRs at various levels: R&S’s *channel* and *acoustic* levels correspond to problems grounding utterance parameters, their *lexical* and *reference* levels to word/phrase parameters, and *intentional*

level at least partly to *max-qud* parameters.* But it goes further, allowing association of specific CR forms with particular types of source and grounding problem. In contrast to plan-based approaches, no inference about plans or intentions behind speech acts is required; and while coercion operations can perhaps be seen as a form of reasoning about context, they are highly constrained and far from general inference.

There is interesting overlap with other work in context-based interpretation: for example, (van der Sandt, 1992)'s view of *presupposition as anaphora*, (Hobbs et al., 1993)'s *interpretation as abduction*, and (Stone, 2003a)'s *pragmatic interpretations*. In all cases, given elements such as definites and names project elements that must be found in context to fully interpret an utterance. The difference here is the focus on clarification: contextual parameters are those that can be clarified. This has inspired the radical abstraction here of e.g. noun and verb predicates, as they are common CR sources; the aforementioned views might lead us to expect (respectively) presupposed events, nominal compound relations, and speaker's intentions to be potential sources – whether this is so remains for further research.

Corrections A possible extension is corrections, which are closely related to CRs; both concern the content or form of a previous utterance, and both can be expressed by similar means (for corrections, primarily the fragment and non-reprise forms (PGH)). They are much less common than CRs in everyday human-human dialogue (about 40 times less (PGH)), but can be important in task-oriented computer dialogues, particularly with speech recognition errors (Lemon and Gruenstein, 2004):*

- (40) Frances: You know Amy?
 Ben: Yeah.
 Frances: Do you reckon that er is, her sister? **Her brother I mean?**
 Ben: Amy? <unclear> Mm.
- (41) User: Fly to the base.
 System: OK. Flying to the tower.
 User: (No,) (not the tower,) **the base**.

(Fernández and Ginzburg, 2002) suggest an analysis directly analogous to CRs, essentially via the *parameter identification* coercion mechanism (definition (24)) within a declarative (rather than interrogative) fragment. For self-corrections such as example (40), this seems reasonable. However, (Gruenstein, 2002) points out that an other-correction such as example (41) needs a paraphrase “*It is **the base** that you are to fly to*”, rather than “*It is **the base** that **you’re telling me** you are to fly to*”, and proposes a correspondingly modified *parameter correction* coercion operation. Both

* Their syntactic parsing level is not covered – as explained in section 4, this is intentional: neither PGH nor R&S found examples that required it.

* Example (41) adapted from (Lemon and Gruenstein, 2004)'s example 5.1.

approaches could be implemented directly within the current system, although care must be taken to disambiguate between these two readings, and between corrections and CRs.

Speech The CLARIE prototype could benefit from extension in many ways: e.g. a larger lexicon, a more realistic domain, and a speech interface. Some of these exist in versions of the TrindiKit-based systems GoDiS and IBiS, and transfer of modules should be feasible in most cases. However, adding speech recognition is a challenge, particularly as regards the interaction of a standard recogniser with the proposed treatment of unknown words: standard speech recognisers make best guesses at known words, rather than hypothesise unknown ones. However, a dual-recogniser approach may be feasible (Hockey et al., 2002; Gorrell, 2003). There are other interesting issues too, for example the use of low recogniser confidence scores to prompt lexical CRs (Larsson, 2002; Gabsdil, 2003), and the use of intonation (pitch contours) for disambiguation. Reprise questions may be distinguishable from statements (Srinivasan and Massaro, 2003), reprise gaps from reprise fragments, and some CR readings from others (R&S show that boundary tone can distinguish lexical from content readings).

Performance & Scaling Up The current CLARIE system is only a prototype with a toy domain and lexicon, intended to show that this general approach to CRs can be implemented in a principled way. On examples such as those in section 5, performance seems reasonable (combined utterance parsing and grounding times below 3s on a standard Windows laptop, 1s on a dual-processor Linux machine). But scaling up to more complex domains with more ambiguity will require more attention to be paid to disambiguating between possible source subconstituents, and CRs from non-CRs. And while the current manually defined rules to disambiguate CR forms and readings are based on corpus evidence, they may vary with domain, language and coverage. As these problems involve the interaction of many complex constraints, they may require an automatic (e.g. machine-learning) approach. Progress has been made into automatic classification and identification of fragment types and antecedents by (Fernández et al., 2004b; Schlangen, 2005); integration of that with the CR approach here and with the CLARIE system itself is in progress.

Conclusions This paper sets out a general linguistic treatment of clarification covering all the most common CR forms and readings, over a wide range of source types; and shows how it can allow a prototype dialogue system to use them to clarify unknown reference and content, and allow users to do the same. Importantly, this is achieved without having to model the user's intentions, or use heavyweight inference about utterances or their interrelation.

User CRs are assigned straightforward (although ambiguous and heavily contextually dependent) representations; and then interpreted by grounding their abstracted parameters in context. Problems with this grounding process, and with particular abstracted parameters, lead to system CRs. CRs are not treated in a significantly different way from other utterances: they are parsed by the same grammar and given a standard interrogative interpretation as *ask* moves which raise new questions for discussion – it is just that these questions concern other utterances.

References

- Barwise, J. and R. Cooper: 1981, ‘Generalized Quantifiers and Natural Language’. *Linguistics and Philosophy* **4**, 159–219.
- Burnard, L.: 2000, *Reference Guide for the British National Corpus (World Edition)*. Oxford University Computing Services.
- Clark, H. H.: 1996, *Using Language*. Cambridge University Press.
- Dalrymple, M., S. M. Shieber, and F. C. N. Pereira: 1991, ‘Ellipsis and Higher-Order Unification’. *Linguistics and Philosophy* **14**(4), 399–452.
- Dusan, S. and J. Flanagan: 2002, ‘Adaptive Dialog Based Upon Multimodal Language Acquisition’. In: *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces*. Pittsburgh.
- Fernández, R. and J. Ginzburg: 2002, ‘Non-Sentential Utterances in Dialogue: A Corpus-Based Study’. In: *Proceedings of the 3rd SIGdial Workshop on Discourse and Dialogue*. Philadelphia, pp. 15–26.
- Fernández, R., J. Ginzburg, H. Gregory, and S. Lappin: 2004a, ‘SHARDS: Fragment Resolution in Dialogue’. In: H. Bunt and R. Muskens (eds.): *Computing Meaning*, Vol. 3. Kluwer Academic Publishers. To appear.
- Fernández, R., J. Ginzburg, and S. Lappin: 2004b, ‘Classifying Ellipsis in Dialogue: A Machine Learning Approach’. In: *Proceedings of the 20th International Conference on Computational Linguistics (COLING)*. Geneva, pp. 240–246.
- Fletcher, C.: 1994, ‘Levels of Representation in Memory for Discourse’. In: M. Gernsbacher (ed.): *Handbook of Psycholinguistics*. Academic Press, pp. 589–607.
- Gabsdil, M.: 2003, ‘Clarification in Spoken Dialogue Systems’. In: *Proceedings of the AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*. Stanford, pp. 28–35.
- Ginzburg, J.: forthcoming, *A Semantics for Interaction in Dialogue*. CSLI Publications. Draft chapters available from: <http://www.dcs.kcl.ac.uk/staff/ginzburg>.
- Ginzburg, J. and R. Cooper: 2004, ‘Clarification, Ellipsis, and the Nature of Contextual Updates in Dialogue’. *Linguistics and Philosophy* **27**(3), 297–365.
- Ginzburg, J. and I. Sag: 2000, *Interrogative Investigations: the Form, Meaning and Use of English Interrogatives*, No. 123 in CSLI Lecture Notes. CSLI Publications.
- Ginzburg, J., I. Sag, and M. Purver: 2003, ‘Integrating Conversational Move Types in the Grammar of Conversation’. In: P. Kühnlein, H. Rieser, and H. Zeevat (eds.): *Perspectives on Dialogue in the New Millennium*, Vol. 114 of *Pragmatics and Beyond New Series*. John Benjamins, pp. 25–42.
- Gorrell, G.: 2003, ‘Using Statistical Language Modelling to Identify New Vocabulary in a Grammar-Based Speech Recognition System’. In: *Proceedings of the 8th European Conference on Speech Communication and Technology (Eurospeech)*. Geneva, pp. 2729–2732.

- Gruenstein, A.: 2002, 'English Corrective Fragments: Syntactic and Semantic Considerations'. Unpublished ms., Stanford University.
- Haddock, N.: 1987, 'Incremental Interpretation and Combinatory Categorical Grammar'. In: *Proceedings of the 10th International Joint Conference on Artificial Intelligence*. Milan, pp. 661–663.
- Healey, P., M. Purver, J. King, J. Ginzburg, and G. Mills: 2003, 'Experimenting with Clarification in Dialogue'. In: *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. Boston, pp. 539–544.
- Heeman, P. and G. Hirst: 1995, 'Collaborating on Referring Expressions'. *Computational Linguistics* **21**(3), 351–382.
- Hobbs, J., M. Stickel, D. Appelt, and P. Martin: 1993, 'Interpretation as Abduction'. *Artificial Intelligence* **63**(1–2), 69–142.
- Hockey, B. A., J. Dowding, G. Aist, and J. Hieronymus: 2002, 'Targeted Help and Dialogue about Plans'. In: *ACL-02 Companion Volume to the Proceedings of the Conference*. Philadelphia, pp. 100–101.
- Hockey, B. A., D. Rossen-Knill, B. Spejewski, M. Stone, and S. Isard: 1997, 'Can You Predict Answers to Yes/No Questions? Yes, No and Stuff'. In: *Proceedings of Eurospeech '97*. Rhodes, pp. 2267–2270.
- Kaplan, D.: 1989, 'Demonstratives: an Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and Other Indexicals'. In: J. Almog, J. Perry, and H. Wettstein (eds.): *Themes from Kaplan*. Oxford University Press, pp. 481–614.
- Kievit, L., P. Piwek, R.-J. Beun, and H. Bunt: 2001, 'Multimodal Cooperative Resolution of Referential Expressions in the DenK System'. In: H. Bunt and R.-J. Beun (eds.): *Cooperative Multimodal Communication: 2nd International Conference*, Vol. 2155 of *Lecture Notes in Computer Science*. Springer, pp. 197–214.
- Larsson, S.: 2002, 'Issue-based Dialogue Management'. Ph.D. thesis, Göteborg University. Also published as Gothenburg Monographs in Linguistics 21.
- Larsson, S., A. Berman, L. Grönqvist, and F. Kronlid: 2002, 'TrindiKit 3.0 Manual'. In: *SIRIDUS deliverable 6.4*.
- Larsson, S., P. Ljunglöf, R. Cooper, E. Engdahl, and S. Ericsson: 2000, 'GoDiS - An Accommodating Dialogue System'. In: *Proceedings of ANLP/NAACL-2000 Workshop on Conversational Systems*. Seattle, pp. 7–10.
- Lemon, O. and A. Gruenstein: 2004, 'Multithreaded Context for Robust Conversational Interfaces: Context-Sensitive Speech Recognition and Interpretation of Corrective Fragments'. *ACM Transactions on Computer-Human Interaction (ACM TOCHI)* **11**(3), 241–267.
- Litman, D. and J. Allen: 1987, 'A Plan Recognition Model for Subdialogues in Conversations'. *Cognitive Science* **11**(2), 163–200.
- Mellish, C.: 1985, *Computer Interpretation of Natural Language Descriptions*. Ellis Horwood.
- Montague, R.: 1974, 'Pragmatics'. In: R. Thomason (ed.): *Formal Philosophy: Selected Papers of Richard Montague*. Yale University Press.
- Pickering, M. and S. Garrod: 2004, 'Toward a Mechanistic Psychology of Dialogue'. *Behavioral and Brain Sciences* **27**, 169–226.
- Purver, M.: 2004, 'The Theory and Use of Clarification Requests in Dialogue'. Ph.D. thesis, University of London.
- Purver, M. and J. Ginzburg: 2004, 'Clarifying Noun Phrase Semantics'. *Journal of Semantics* **21**(3), 283–339.
- Purver, M., J. Ginzburg, and P. Healey: 2003a, 'On the Means for Clarification in Dialogue'. In: R. Smith and J. van Kuppevelt (eds.): *Current and New Directions in Discourse & Dialogue*. Kluwer Academic Publishers, pp. 235–255.

- Purver, M., P. Healey, J. King, J. Ginzburg, and G. Mills: 2003b, 'Answering Clarification Questions'. In: *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue*. Sapporo, pp. 23–33.
- Reeves, B. and C. Nass: 1996, *The Media Equation: How People Treat Computers, Television and New Media like Real People and Places*. Cambridge University Press.
- Rodríguez, K. and D. Schlangen: 2004, 'Form, Intonation and Function of Clarification Requests in German Task-Oriented Spoken Dialogues'. In: *Proceedings of the 8th Workshop on the Semantics and Pragmatics of Dialogue (Catalog)*. Barcelona, pp. 101–108.
- Schlangen, D.: 2003, 'A Coherence-Based Approach to the Interpretation of Non-Sentential Utterances in Dialogue'. Ph.D. thesis, University of Edinburgh.
- Schlangen, D.: 2004, 'Causes and Strategies for Requesting Clarification in Dialogue'. In: *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue*. Boston, pp. 136–143.
- Schlangen, D.: 2005, 'Towards Finding and Fixing Fragments: Using Machine Learning to Identify Non-Sentential Utterances and Their Antecedents in Multi-Party Dialogue'. In: *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*. Ann Arbor, pp. 247–254.
- SFB-360: 2000, 'Bielefeld Corpus'. <http://www.sfb360.uni-bielefeld.de>.
- Srinivasan, R. and D. Massaro: 2003, 'Perceiving Prosody from the Face and Voice: Distinguishing Statements from Echoic Questions in English'. *Language and Speech* **46**(1), 1–22.
- Stone, M.: 2003a, 'Linguistic Representation and Gricean Inference'. In: *Proceedings of the 5th International Workshop on Computational Semantics (IWCS-5)*. Tilburg, pp. 5–21.
- Stone, M.: 2003b, 'Specifying Generation of Referring Expressions by Example'. In: *Proceedings of the AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*. Stanford, pp. 133–140.
- Traum, D.: 2003, 'Semantics and Pragmatics of Questions and Answers for Dialogue Agents'. In: *Proceedings of the 5th International Workshop on Computational Semantics (IWCS-5)*. Tilburg, pp. 380–394.
- van der Sandt, R.: 1992, 'Presupposition Projection as Anaphora Resolution'. *Journal of Semantics* **9**, 333–377.