Chapter 1

ON THE MEANS FOR CLARIFICATION IN DIALOGUE

Matthew Purver
Jonathan Ginzburg
Department of Computer Science
King's College London
Strand, London WC2R 2LS, UK
{matthew.purver, jonathan.ginzburg}@kcl.ac.uk

Patrick Healey
Department of Computer Science
Queen Mary, University of London
Mile End Road, London E1 4NS, UK
ph@dcs.qmul.ac.uk

Abstract   The ability to request clarification of utterances is a vital part of the communicative process. In this paper we discuss the range of possible forms for clarification requests, together with the range of readings they can convey. We present the results of corpus analysis which show a correlation between certain forms and possible readings, together with some indication of maximum likely distance between request and the utterance being clarified. We then explain the implications of these results for a possible HPSG analysis of clarification requests and for an ongoing implementation of a clarification-capable dialogue system.

Keywords: Clarification, dialogue, HPSG

1.   Introduction

Clarification requests (CRs) are common in human conversation. They can take various *forms* and can be intended by the speaker making the request (the CR *initiator*) to request various types of clarification information (i.e. they can have various *readings*), but have in common the fact that they are in some sense *utterance-anaphoric* – they concern the content or form of a previous utterance that has failed to be fully comprehended by the initiator.

D R A F T   October 31, 2002, 12:35pm   D R A F T

It is not usual for computer dialogue systems do be able to process CRs produced by the user. One can see how important this might be in a negotiative dialogue by considering the following imagined exchange, which gives some possible alternative responses to a CR initiated by the caller:

(1)

System:	Would you like to travel via Paris or Amsterdam?
Caller:	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 caller's move as a CR, might be regarded as useful to the caller: response (d), which incorrectly interprets it as an answer to the system's question, would not be acceptable under any circumstances. Which of (a)–(c) is preferred will depend on the reading intended. As a first step towards a full theory of CR interpretation, we therefore believe it is important to have information about which readings are available via which forms.

Previous studies have examined some individual CR forms and given possible analyses for these forms. In this paper we describe an attempt to exhaustively categorise CR forms and readings based on corpus work, and discuss the implications of our results for further analysis.

The analyses so far proposed require all information from a previous utterance to be retained in memory (not only propositional content but syntax and phonology). The retention of such a large amount of information indefinitely poses obvious problems for any implementation with finite resources, and seems at odds with some results from work in psycholinguistics: studies such as (Sachs, 1967; van Dijk and Kintsch, 1983) have argued that surface information such as syntax is retained only in the short term (see (Fletcher, 1994) for an overview). Our corpus work has therefore had the additional aim of identification of the maximum distance between a CR and the utterance being clarified (the *source* utterance).

In this section we give a brief overview of CR forms identified in previous work, together with the analyses proposed and the readings that these analyses give rise to. In sections 2 and 3 we list the possible CR forms and readings that we have identified from corpus analysis. In section 4 we describe this analysis and give detailed results, including a discussion of apparent correlations between certain forms and readings and of maximum observed CR-source separation (CSS) distance. Finally, in section 5 we discuss the implications of our findings for an intended dialogue system implementation.

1.1 Previous Work

(Ginzburg and Sag, 2000) (hereafter G&S) discuss *reprise interrogatives*, which they further classify into *echo* questions (those "resulting from mishearing a previous speech act" – see B's question in example (2)) and *reference* questions (those which "ask for clarification of the reference of some element in the immediately prior utterance" – see example (3)).

(2) | A: | Did Jill phone?
     | B: | Did JILL phone?

(3) | A: | Did Jill phone?
     | B: | Did WHO phone?

They argue that the content of both readings "contains as a constituent the illocutionary force of the (previous) utterance" being reprised. In other words, B's utterances in the examples above both involve querying some feature of A's query. They might be paraphrased *"Are you asking whether Jill phoned?"* and *"For which person are you asking whether that person phoned?"*, respectively.

They therefore offer a syntactic and semantic analysis which covers both readings: the reprise is analysed syntactically as an *in-situ* interrogative, and semantically as a question which takes as its propositional content the perceived content of the previous utterance being clarified. As conversational move type (CMT) is integrated into utterance content by their HPSG grammar (see (Ginzburg et al., 2001b)) this straightforwardly gives rise to a reading along the lines of *"For which X are you asking/asserting/(etc.) Y about X?"*. They give a full derivation for this reading based on the KOS dialogue context framework (Ginzburg, 1996; Bohlin (Ljunglöf) et al., 1999).

This analysis is then extended to two elliptical forms: *reprise sluices* and *elliptical literal reprises*. Sluices are elliptical wh-constructions (see (Ross, 1969)) – short wh-questions which receive a "sentential" interpretation, in this case an interpretation as a reprise question, as shown in example (4):

(4) | A: | Did Jill phone?
     | B: | WHO?
     | | (non-elliptical equivalent: Did WHO phone?)

Elliptical literal reprises are short polar questions – bare fragments which receive an interpretation as a polar reprise question:

(5) | A: | Did Jill phone?
     | B: | JILL?
     | | (non-elliptical equivalent: Did JILL phone?)

Resolution of these elliptical forms is achieved by allowing a conversational participant to coerce a clarification question onto the list of questions under discussion (QUD) in the current dialogue context. This allows ellipsis resolution

in the manner of of (Ginzburg et al., 2001a) to give essentially the same reading as reprise questions.

(Ginzburg and Cooper, 2001) (hereafter G&C) give more detailed analysis for the bare fragment form (therein described as *clarification ellipsis*) and also give a further reading for this form. They call this reading the *constituent* reading to distinguish it from the *clausal* reading described above. This constituent reading involves querying the content of a constituent which the CR initiator has been unable to ground in context (see (Traum, 1994; Clark, 1996)), and is along the lines of *"What/who/(etc.) is the reference of your utterance X?"*.

A possible *lexical identification* reading is also discussed, but no analysis given. They also raise the issue of whether these specific readings really exist or could be subsumed by a single vague reading, but give evidence that this is not the case: they cite examples of CR misunderstanding leading to repeated attempts to elicit the desired clarificational information, showing that a specific reading was intended; they also point out that some readings involve different parallelism conditions. As will be discussed in detail below, the results of the work described here also indicate that particular forms may be restricted to particular sets of specific readings.

2. Clarification Forms

The following forms have been identified as possible means for CRs. While we cannot claim that this list is exhaustive, a markup scheme based on these forms has been shown to cover the CRs encountered in a corpus of dialogue, as detailed in section 4 below. In this section we list the forms identified, and illustrate them with examples. All examples have been taken from the British National Corpus (BNC).

2.1 Non-Reprise Clarifications

Unsurprisingly, speakers have recourse to a non-reprise[1] form of clarification. In this form, the nature of the information being requested by the CR initiator is spelt out for the addressee. Utterances of this type thus often contain phrases such as *"do you mean..."*, *"did you say..."*, as can be seen in examples (6) and (7).

(6)[2]
Cassie:	You did get off with him?
Catherine:	Twice, but it was totally non-existent kissing so
Cassie:	**What do you mean?**
Catherine:	I was sort of falling asleep.

(7)[3]
Leon:	Erm, your orgy is a food orgy.
Unknown:	**What did you say?**
Leon:	Your type of orgy is a food orgy.

2.2   Reprise Sentences

Speakers can form a CR by echoing or repeating[4] a previous utterance in full, as shown in example (8). This form corresponds to G&S's *reprise interrogative*.

(8)[5]
Orgady:	I spoke to him on Wednesday, I phoned him.
Obina:	**You phoned him?**
Orgady:	Phoned him.

This form appears to be divisible into two sub-categories, *literal* (as in example (8) above) and *wh-substituted* reprise sentences, as illustrated by example (9).

(9)[6]
Unknown:	He's anal retentive, that's what it is.
Kath:	**He's what?**
Unknown:	Anal retentive.

2.3   Reprise Sluices

This form is an elliptical wh-construction as already discussed above and described by G&S.

(10)[7]
Sarah:	Leon, Leon, sorry she's taken.
Leon:	**Who?**
Sarah:	Cath Long, she's spoken for.

There may be a continuum of forms between *wh-substituted reprise sentences* and *reprise sluices*. Consider the following exchange (11):

(11)[8]
| Richard: | I'm opening my own business so I need a lot of money |
| Anon 5: | **Opening what?** |

This form seems to fall between the full wh-substituted reprise sentence *"You're opening (your own) what?"* and the simple reprise sluice *"(Your own) what?"*. The actual form employed in this case appears closer to the sluice and was classified as such.[9]

2.4   Reprise Fragments

This elliptical bare fragment form corresponds to that described as *elliptical literal reprise* by G&S and *clarification ellipsis* by G&C.

$(12)^{10}$

Lara:	There's only two people in the class.
Matthew:	**Two people?**
Unknown:	For cookery, yeah.

A similar form was also identified in which the bare fragment is preceded by a wh-question word:

$(13)^{11}$

Ben:	No, ever, everything we say she laughs at.
Frances:	**Who Emma?**
Ben:	Oh yeah.

As these examples appeared to be interchangeable with the plain fragment alternative (in example (13), *"Emma?"*), they were not distinguished from fragments in our classification scheme.

2.5   Gaps

The *gap* form differs from the reprise forms described above in that it does not involve a reprise component corresponding to the component being clarified. Instead, it consists of a reprise of (a part of) the utterance immediately preceding this component – see example (14).

$(14)^{12}$

Laura:	Can I have some toast please?
Jan:	**Some?**
Laura:	Toast

Our intuition is that this form is intonationally distinct from the reprise fragment form that it might be taken to resemble. This appears to be backed up by the fact that no misunderstandings of gap-CRs were discovered during our corpus analysis.

2.6   Gap Fillers

The *filler* form is used by a speaker to fill a gap left by a previous incomplete utterance. Its use therefore appears to be restricted to such contexts, either because a previous speaker has left an utterance "hanging" (as in example (15)) or because the CR initiator interrupts.

$(15)^{13}$

Sandy:	if, if you try and do enchiladas or
Katriane:	Mhm.
Sandy:	erm
Katriane:	**Tacos?**
Sandy:	tacos.

2.7 Conventional

A *conventional* form is available which appears to indicate a complete breakdown in communication. This takes a number of seemingly conventionalised forms such as *"What?"*, *"Pardon?"*, *"Sorry?"*, *"Eh?"*:

(16)[14]

	Anon 2:	Gone to the cinema tonight or summat.
	Kitty:	**Eh?**
	Anon 2:	Gone to the cinema

3. Clarification Readings

This section presents the readings that have been identified, together with examples. We follow G&C's proposed *clausal*/*constituent*/*lexical* split, with an added reading for *corrections*.

3.1 Clausal

The *clausal* reading takes as the basis for its content the *content of the conversational move* made by the utterance being clarified.

This reading corresponds roughly to *"Are you asking/asserting that X?"*, or *"For which X are you asking/asserting that X?"*. It follows that the source utterance must have been partially grounded by the CR initiator, at least to the extent of understanding the move being made.

An attribute-value matrix (AVM) skeleton for the semantic content of an HPSG sign corresponding to this reading (according to G&C's analysis) is shown below as AVM [1]. It represents a question[15], the propositional content of which is the conversational move made by the source utterance (shown here as being of type *illoc(utionary)-rel(ation)* – possible subtypes include *assert, ask*) together with the message associated with that move (e.g. the proposition being asserted). The parameter set being queried can be either a constituent of that message (as would be the case in a sluice or wh-substituted form, where the CR question is the wh-question *"For which X are you asserting . . . "*) or empty (as would be the case in a fragment or literal reprise form, where the CR question is the polar question *"Are you asserting . . . "*).

$$
[1] \quad
\begin{bmatrix}
question \\
\text{PARAMS} \quad \left\{ \boxed{2} \right\} \text{or} \left\{ \right\} \\
\text{PROP} \mid \text{SOA} \quad
\begin{bmatrix}
illoc\text{-}rel \\
\text{UTTR} \quad \boxed{1} \\
\text{MSG-ARG} \quad \left[\ldots \boxed{2} \ldots \right]
\end{bmatrix}
\end{bmatrix}
$$

3.2   Constituent

Another possible reading is a *constituent* reading whereby the content of *a constituent* of the previous utterance is being clarified.

This reading corresponds roughly to *"What/who is X?"* or *"What/who do you mean by X?"*, as shown in AVM [2], a description of the content that would be given by G&C's analysis. This shows a question whose propositional content is the relation between a sign (a constituent of the source utterance), its speaker, and the intended semantic content. The abstracted parameter is the content.

$$
[2] \quad
\begin{bmatrix}
\textit{question} & \\
\text{PARAMS} & \left\{ \boxed{3} \right\} \\
\text{PROP} \mid \text{SOA} &
\begin{bmatrix}
\textit{spkr-meaning-rel} \\
\text{AGENT} \quad \boxed{1} \\
\text{SIGN} \quad \boxed{2} \\
\text{CONT} \quad \boxed{3}
\end{bmatrix}
\end{bmatrix}
$$

3.3   Lexical

Another possibility appears to be a *lexical* reading. This is closely related to the clausal reading, but is distinguished from it in that the *surface form* of the utterance is being clarified, rather than the content of the conversational move.

This reading therefore takes the form *"Did you utter X?"* or *"What did you utter?"*. The CR initiator is attempting to identify or confirm a word in the source utterance, rather than a part of the semantic content of the utterance. This poses some interesting questions if a full analysis for this reading is to be integrated into the HPSG framework described above.

3.4   Corrections

The correction reading appears be along the lines of *"Did you intend to utter X (instead of Y)?"*. We do not as yet have a full analysis for this reading.[16]

4.   Corpus Analysis

4.1   Aims and Procedure

Our intention was to investigate the forms and readings for CRs that are present in a corpus of dialogue. For this purpose we used the BNC, which contains a 10 million word sub-corpus of English dialogue transcripts. For this experiment, a sub-portion of the dialogue transcripts was used consisting of c. 150,000 words. To maintain a spread across dialogue domain, region, speaker age etc., this sub-portion was created by taking a 200-speaker-turn section from 59 transcripts.

All CRs within this sub-corpus were identified and tagged, using the markup scheme and decision process described in 4.2 and 4.3 below. The results given here are those produced by an expert, although the process has been repeated by a naive user to check its reliability (this was found to be reasonable – see below).

Initial identification of CRs was performed using SCoRE (Purver, 2001), a search engine developed specifically for this purpose (in particular, to allow searches for repeated words between speaker turns, and to display dialogue in an intuitive manner). However, in order to ensure that all clarificational phenomena were captured, the final search and markup were performed manually.

4.2 Markup Scheme

The markup scheme used evolved during the markup process as new CR mechanisms were identified, and the final scheme was as described here. A multi-layered approach was taken, along the lines of the DAMSL dialogue act markup scheme (Allen and Core, 1997) – this allowed sentences to be marked independently for three attributes: *form*, *reading* and *source*.

The *form* and *reading* attributes had finite sets of possible values. The possible values were as described in sections 2 and 3, plus an extra catch-all category *other* to deal with any otherwise uncategorisable phenomena. The *source* attribute could take any integer value and was set to the number of the sentence that was being clarified (according to the BNC sentence-numbering scheme).

4.3 Decision Process

Following the methods described in (Allen and Core, 1997), binary decision trees were designed to guide the classification process. The trees are designed so that a naive user can follow them. Trees were produced for for classification of CR form and reading and for determination of CR source, and are shown in appendix 1.A.

In the (common) case of ambiguity of reading, the response(s) of other dialogue participants were examined to determine which reading was chosen by them. The ensuing reaction of the CR initiator was then used to judge whether this interpretation was acceptable. If the CR initiator gave no reaction, the reading was assumed to have been acceptable. The following example (17) shows a case where the other participant's initial (clausal) reading was incorrect (the

initiator is not satisfied), as a constituent reading was required. In such cases, both CRs were marked as constituent.

(17)[17]

George:	you always had er er say every foot he had with a piece of spunyarn in the wire
Anon 1:	**Spunyarn?**
George:	Spunyarn, yes
Anon 1:	**What's spunyarn?**
George:	Well that's like er tarred rope

In example (18), however, the other participant's clausal interpretation provokes no further reaction from the CR initiator, and is taken to be correct:

(18)[18]

Anon 1:	you see the behind of Taz
Selassie:	**Tazmania?**
Anon 1:	Yeah.
Selassie:	Oh this is so rubbish man.

In order to facilitate this process in the case of CRs near the beginning or end of the 200-turn section being marked, an additional 10 turns of backward and forward context were shown (but not themselves marked up).

In the case of ambiguity as to which sentence was being clarified, the most recent one was taken as the source.

Reliability. The markup process was repeated (after an interval of several months) by the expert user, and was also performed by a naive user. The results were then compared both in terms of raw agreement and the kappa statistic (Carletta, 1996), to assess reliability. The kappa statistic gives an indication of the level of agreement above that level which would be expected randomly.

A kappa figure $\kappa \geq 80\%$ is generally considered to indicate good reliability, with $\kappa \geq 67\%$ being good enough to draw tentative conclusions. The results are shown in table 1.1 below: CR source figures can be seen to be good, with figures for form & reading above or close to the 80% level for both expert and naive users. We believe that these levels are acceptable for the rather general conclusions we draw about form & reading distribution in this paper.

Table 1.1. Markup Reliability

	Raw (expert) (%)	Raw (naive) (%)	Kappa (expert) (%)	Kappa (naive) (%)
Form	90	83	88	78
Reading	85	84	77	75
Source	95	92	90	83

Examination of confusion matrices has shown that in the case of form, confusion is shared roughly equally between genuine ambiguity (e.g. conventional

vs. sluice "what?", fragment vs. gap) and uncertainty of the classification scheme (e.g. the continuum of forms between full and elliptical reprise mentioned in section 2.3 above). Confusion of reading appeared to be due to genuine ambiguity in the corpus as presented.

We believe that much of the genuine ambiguity could be resolved by use of an audio corpus (or text corpus containing intonational data). An improved classification scheme and instructions could help with the uncertainty, in particular by conflating certain categories (if subsequent research shows that this is desirable).

4.4   Results

The BNC's SGML markup scheme (see (Burnard, 2000) for details) allows sub-corpora to be easily identified according to domain. This allowed us to collate results both over all dialogue domains[19], and restricted to dialogue identified as *demographic* (non-context-governed).

The distribution of CRs by form and reading are shown in full in table 1.2 (all dialogue domains) and table 1.3 (demographic only). The distributions are presented as percentages of all CRs found. This allows us to see the proportion made up by each form and each reading, together with any correlations between form and reading, as discussed in full below. Distributions are similar over both sets, indicating that corpus size is large enough to give repeatable results.

Separation between CR and source sentence was calculated in terms both of sentences and speaker turns (both are marked in the BNC). According to the BNC markup system, one speaker turn can consist of more than one sentence, as expected; it can also consist of zero sentences in cases where the contribution was non-verbal or was unclear to the transcriber. The distributions are presented as number of CRs in tables 1.4 and 1.5, and as cumulative percentages in figures 1.1 and 1.2.

Table 1.2. CR form and type as percentage of CRs – all domains

	Non-Reprise	Literal Reprise	Wh-sub Reprise	Reprise Sluice	Reprise Fragmt	Gap	Gap Filler	Conventional	Other	Total
Clausal	4.3	4.8	1.0	10.7	25.3	0	0	0	0.5	46.6
Constituent	7.7	0	0	0	1.7	0	0	5.3	0	14.5
Lexical	0.7	0	2.6	2.2	0.2	0.5	3.8	24.8	0	34.9
Correction	1.0	0.5	0	0	1.0	0	0	0	0	2.4
Other	0	0	0	0	1.0	0	0	0.5	0	1.4
Total	13.6	5.3	3.6	12.9	29.1	0.5	3.8	30.6	0.5	100.0

Table 1.3. CR form and type as percentage of CRs – demographic portion

	Non-Reprise	Literal Reprise	Wh-sub Reprise	Reprise Sluice	Reprise Fragmt	Gap	Gap Filler	Conventional	Other	Total
Clausal	4.1	4.7	1.0	11.3	24.8	0	0	0	0.5	46.6
Constituent	6.2	0	0	0	1.8	0	0	5.7	0	13.7
Lexical	0.8	0	2.6	2.3	0.3	0.5	3.1	26.1	0	35.7
Correction	1.0	0.5	0	0	1.0	0	0	0	0	2.6
Other	0	0	0	0	0.8	0	0	0.5	0	1.3
Total	12.1	5.2	3.6	13.7	28.7	0.5	3.1	32.3	0.5	100.0

Table 1.4. Number of CRs vs. CR-Source Separation Distance (Sentences)

Distance	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
All domains	1	297	35	16	10	3	4	0	2	2	0	0	0	1	0	1
Demographic	0	270	33	16	9	3	4	0	2	2	0	0	0	1	0	1

Table 1.5. Number of CRs vs. CR-Source Separation Distance (Turns)

Distance	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
All domains	11	314	14	16	9	2	1	1	1	0	1	1	0	0	1	0
Demographic	9	287	13	15	9	2	1	1	1	0	1	1	0	0	1	0

Form/Reading Distribution. CRs were found to make up just under 4% of sentences when calculated over the demographic portion, or just under 3% when calculated over all domains. This is a significant proportion, giving support to our claim that processing of CRs is important for a dialogue system.[20]

The most common forms of CR can be seen to be the conventional and reprise fragment forms, with each making up over 25% of CRs. Non-reprise CRs and reprise sluices are also common, each contributing over 10% of CRs. Other forms are all around 5% or less.

Nearly 50% of CRs can be successfully interpreted as having a clausal reading, although both the lexical (about 35%) and constituent (about 15%) readings also make up a significant proportion.

This initially suggests that an automated dialogue system which can deal with fragments, sluices and reprise sentences (the analyses described in section 1), together with conventional and non-reprise CRs, could give reasonable coverage of expected dialogue. Fillers and especially gaps make up only a small proportion.

However, the high proportion of lexical readings suggests that a detailed analysis of this phenomenon will be required.

Coverage. The coverage of the corpus by the forms and readings listed in this paper is good, with only 0.5% of CR readings (2 sentences) and about 1.5% of CR forms (6 sentences) being classified as *other*.

The readings not covered were all expressing surprise, amusement or outrage at a previous utterance (rather than requesting clarification directly), and were all of the reprise fragment or conventional form. Our intuition is that these readings can be treated as clausal readings with a further level of illocutionary force given by use in context.[21]

Of the 2 sentences left unclassified for form, one appears to be an unusual conventional reading, and one an interesting example of a literal reprise of an unuttered but implied sentence.

Form/Reading Correlation. It appears that of the non-conventional reprise forms, only the reprise fragment requires an analysis that gives a constituent reading. Even then, this reading is much less common than the clausal reading, and we intend further investigation into this fact. Sluices and reprise sentences appear always to be satisfactorily interpretable by a clausal or lexical reading.[22]

As few examples of the rarer forms were observed, it would be dangerous to attempt to draw any firm conclusions about the readings they can carry. We can, however, tentatively suggest that the gap and filler forms might only be used with a lexical reading.[23]

One conclusion that can be safely drawn is that many readings are available for some forms (for example, the reprise fragment form which appears to allow

all readings). This implies that disambiguation between readings will be important for a dialogue system, and this is an area we are currently examining. Possibilities for sources of information that could be used for disambiguation include dialogue context and intonation.

CR-Source Separation. The maximum CSS distance observed was 15 sentences. Only one example of this distance was observed, and one example of distance 13 – otherwise all CSS distances were below 10 sentences. It should be noted that the two long-distance cases were both seen in one dialogue which had more than one speaker present (the dialogue was in a classroom situation with many people talking and one speaker attempting to clarify an utterance by the teacher), so may not be entirely representative of the situation expected with an automated dialogue system.

The vast majority (about 80%) of CRs had a CSS distance of one (i.e. were clarifying the immediately preceding sentence – see figure 1.1), and over 96% had a distance of 4 sentences or less.

The pattern in terms of speaker turns is similar, if slightly more pronounced: nearly 85% had a CSS distance of one turn, and over 95% had a distance of 3 turns or less (see figure 1.2). Two differences between the figures for sentences and turns may be worth pointing out. Firstly, a zero distance is more common for turns: self-clarification (usually correction) within a turn is reasonably common – about 3% of examples – whereas it is rare within a sentence, (only one example was found). Secondly, while the number of occurences decreases smoothly with CSS distance (above 0) for sentences, we can see that this is not quite true for turns: a distance of 3 seemed to be more common that a distance of 2. This may be an indication that clarification of another participant's contributions is more common than clarification of one's own (assuming that alternate turn-taking is the norm, odd CSS distances suggest clarification of others, even distances clarification of self).

5.   Conclusions

The taxonomy of readings and forms given in this paper has been shown to cover nearly 99% of CRs within a corpus of dialogue. A full HPSG analysis has been given elsewhere for two of the four readings and four of the eight forms.

Of the remaining readings, we believe that the lexical reading can be treated by an extension of the existing analysis. Corrections will need further research but make up only a small proportion of CRs.

Of the remaining forms, we believe that two (non-reprise and conventional) can be accommodated relatively smoothly within our current HPSG framework. Gaps and fillers, however, present a significant challenge and will be the subject of future research.

Figure 1.1. Percentage of CRs vs. CR-Source Separation Distance (Sentences)
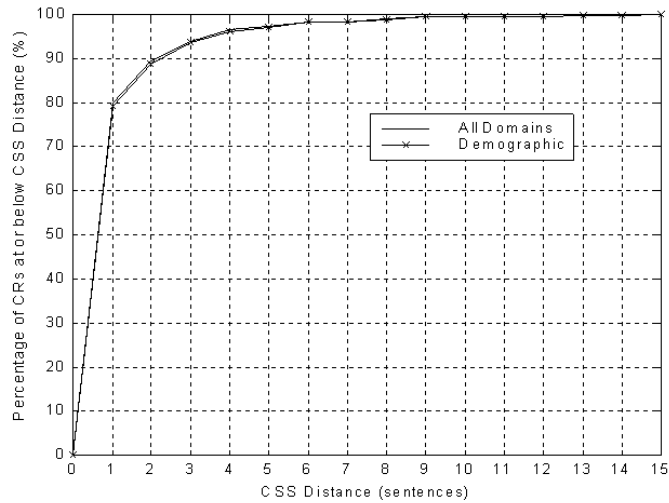

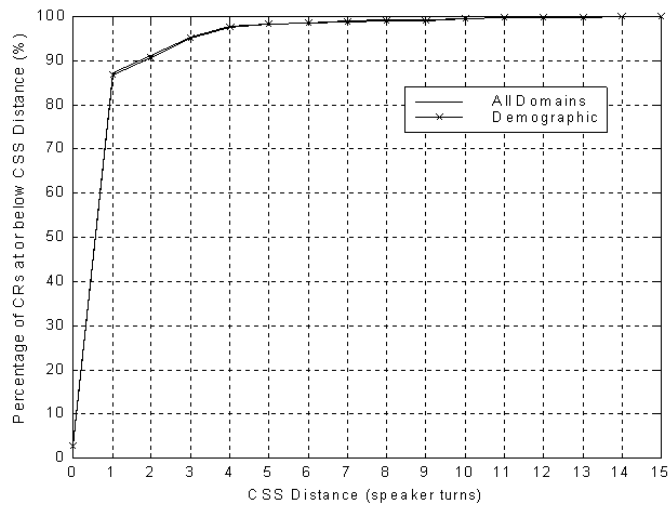*Figure 1.2.* Percentage of CRs vs. CR-Source Separation Distance (Turns)

The measurements of CSS distance show that an utterance record with length of the order of ten sentences would be sufficient to allow a dialogue system to process the vast majority of CRs.

We are in the process of implementing our existing analyses for the CR forms and readings described above within a HPSG/TrindiKit-based dialogue system

which incorporates the ellipsis resolution capability of SHARDS (Ginzburg et al., 2001a) and the dialogue move engine of GoDiS (Larsson et al., 2000). At time of writing, the system can successfully produce both clausal and constituent readings. As a result of the research outlined in this paper, a lexical reading is currently being implemented.

Our results also suggest that investigation into disambiguation of reading, possibly on the basis of dialogue information state and/or intonation, will be required.

Acknowledgments

Appendix: Corpus Markup Decision Trees
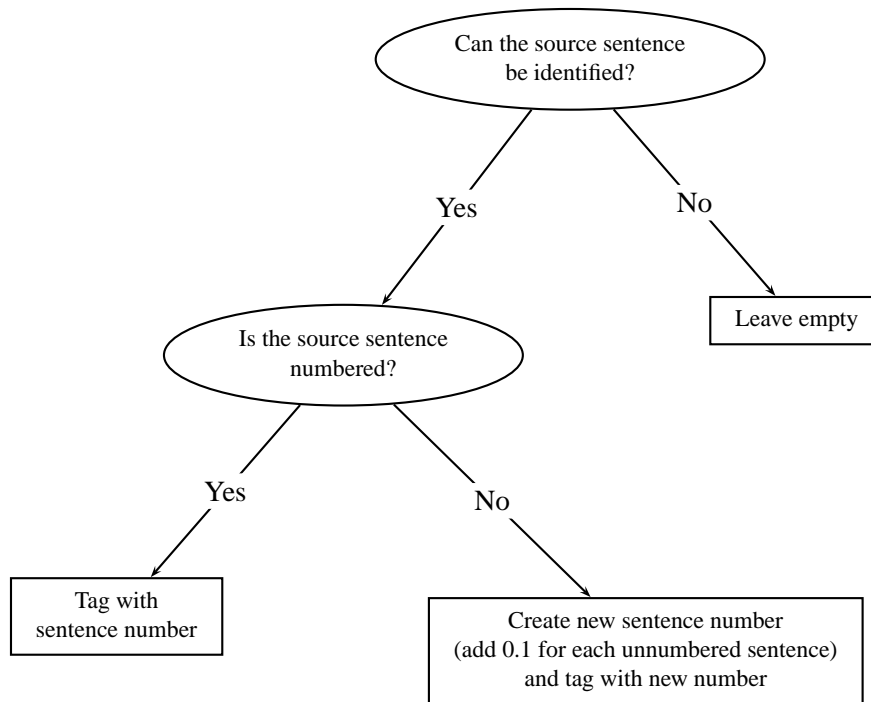
Figure 1.A.1. Decision Tree: CR Source

Does the CR literally specify the nature
of the information being requested?

Yes

No

Tag as
non

Is the CR a conventional phrase
indicating complete incomprehension?

Yes

No

Tag as
wot

Does the CR echo a complete (could stand
in its own right) sentential part of a previous
utterance in order to clarify that part?

Yes

No

Is part of this echoed utterance
replaced by a wh-question word?

Does the CR echo a fragment
of a previous utterance in order
to clarify that fragment?

Yes

No

No

Tag as
sub

Tag as
lit

Yes

Does the CR echo a part
of a previous utterance in order
to clarify the following part?

Is part of this fragment
replaced by a wh-question word?

Yes

No

Yes

No

Tag as
slu

Tag as
frg

Tag as
gap

Does the CR provide a possible part
of an unfinished previous utterance?

Yes

No

Tag as
fil

Tag as
oth

Figure 1.A.2. Decision Tree: CR Form

D R A F T   October 31, 2002, 12:35pm   D R A F T

Can the meaning of the CR be expressed as
"[For which X] are you asking/asserting/. . . X . . . ?"?

Yes → Tag as `cla`

No →

Can the meaning of the CR be expressed as
"[For which X] did you utter X . . . ?"?

Yes → Tag as `lex`

No →

Can the meaning
of the CR be expressed as
"[What] did you mean by X . . . ?"?
or *"[What] is X . . . ?"*?

Yes → Tag as `con`

No →

Can the meaning of the CR be expressed as
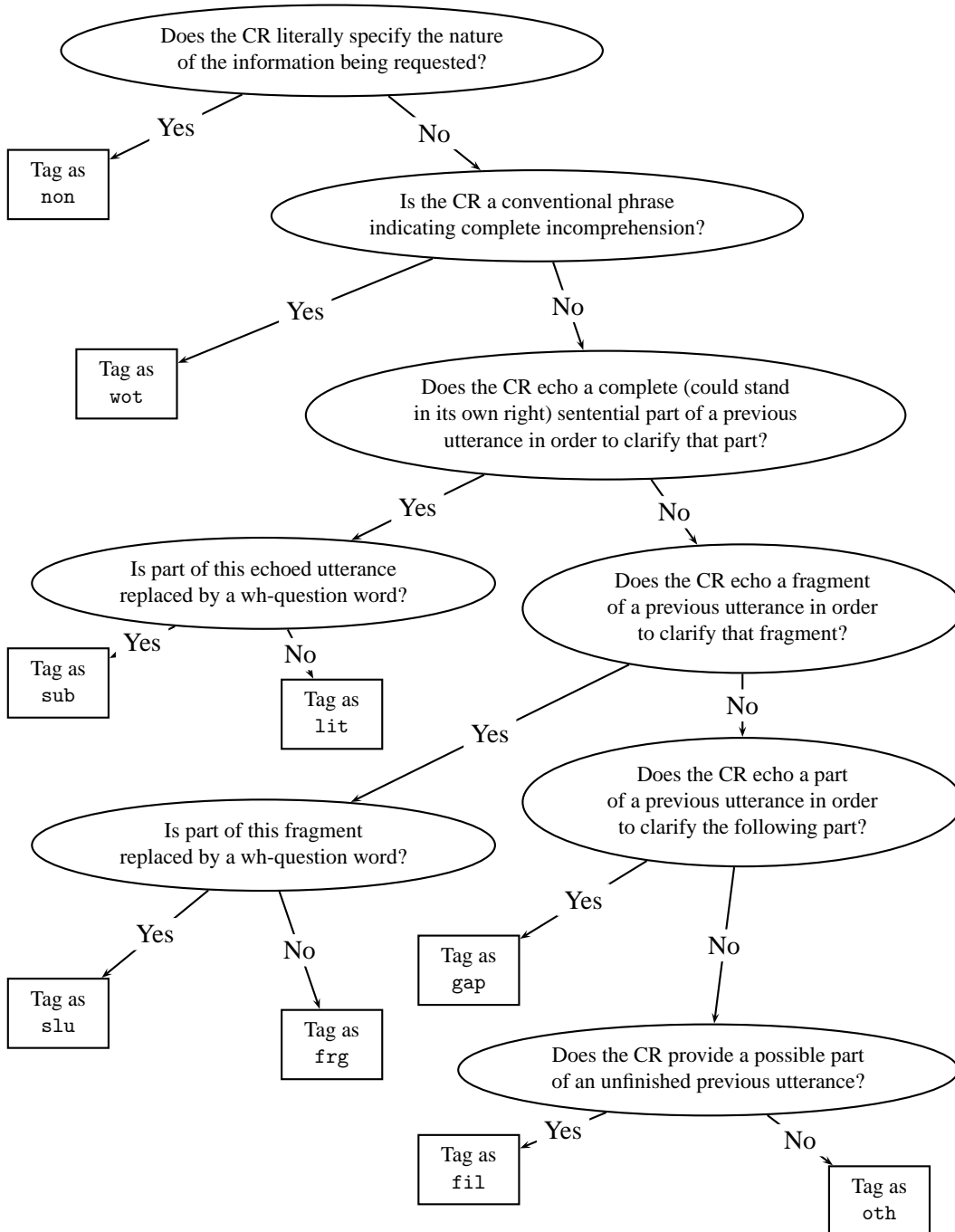*"Did you intend to utter/ask/assert/. . . X (not Y). . . ?"*?
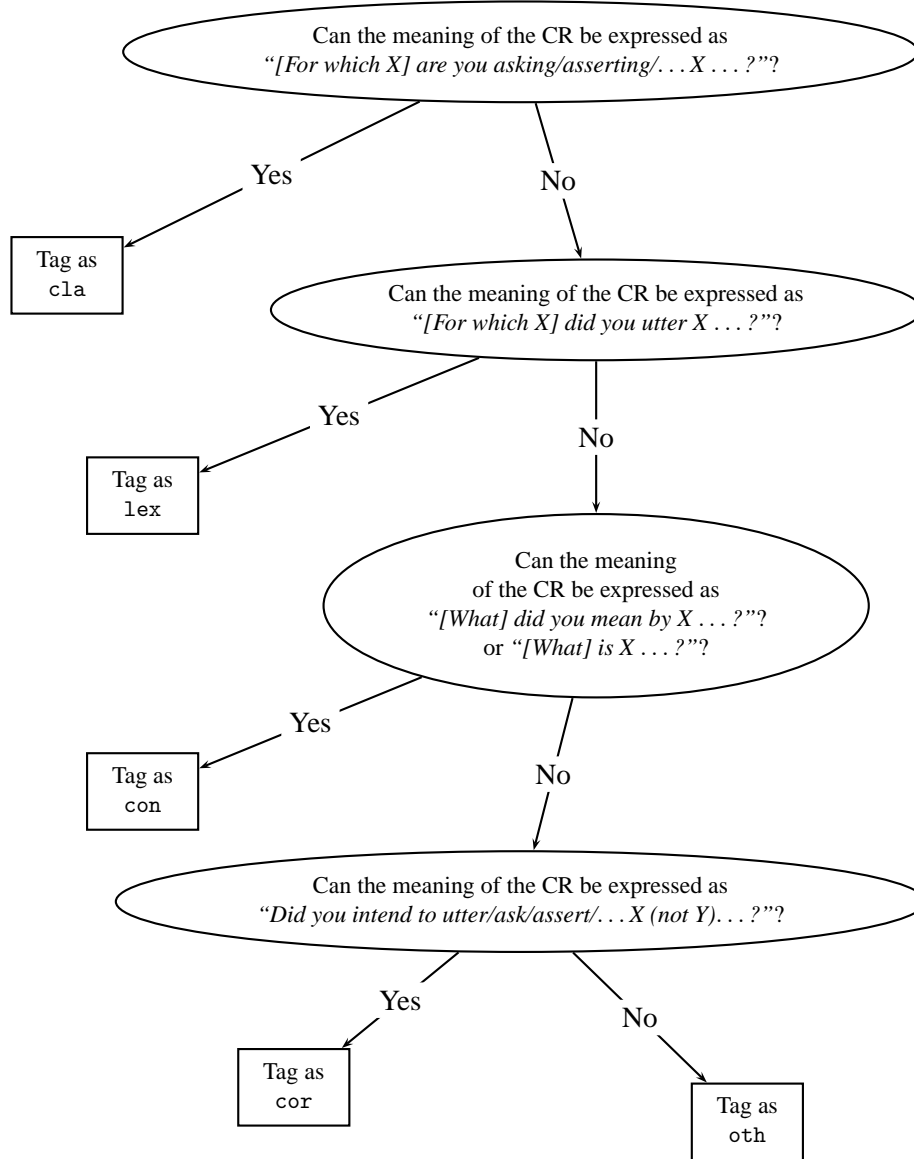
Yes → Tag as `cor`

No → Tag as `oth`

Figure 1.A.3.   Decision Tree: CR Reading

Notes

1. Note that a *non-reprise* sentence need not be *non-elliptical*.
2. BNC file KP4, sentences 521–524
3. BNC file KPL, sentences 524–526
4. Repeats need not be verbatim, due to the possible presence of phenomena such as anaphora and VP ellipsis, as well as changes in indexicals as shown in example (8).
5. BNC file KPW, sentences 463–465
6. BNC file KPH, sentences 412–414
7. BNC file KPL, sentences 347–349
8. BNC file KSV, sentences 363–364
9. While the current exercise has not highlighted it as an issue, we note that a similar continuum might be present between literal reprises and reprise fragments. One approach in the face of this indeterminacy might be to conflate these forms – further analysis of the results given in this paper may indicate whether this is desirable.
10. BNC file KPP, sentences 352–354
11. BNC file KSW, sentences 698–700
12. BNC file KD7, sentences 392–394
13. BNC file KPJ, sentences 555–559
14. BNC file KPK, sentences 580–582
15. We adopt here the version of HPSG developed in G&S, wherein questions are represented as semantic objects comprising a set of parameters (empty for a polar question) and a proposition. This is the feature-structure counterpart of a λ-abstract wherein the parameters are abstracted over the proposition.
16. We suspect that corrections can in fact have clausal, constituent or lexical sub-type, so this may in fact not be a separate reading but a particular usage of those already established. In this case corrections may be covered by the analyses given for other readings above, with a modified QUD coercion operation – see (Ginzburg and Cooper, ming).
17. BNC file H5G, sentences 193–196
18. BNC file KNV, sentences 548–551
19. Domains identified by the BNC as context-governing for dialogue include educational (school classes, lectures) and business (meetings, training sessions) – see (Burnard, 2000) for a full list.
20. Although this proportion is calculated on the basis of human-human dialogue, David Traum has indicated that a similar proportion was observed during the TRAINS experiments.
21. Jens Allwood has brought to our attention the fact that several further readings are possible (for example, a "courtroom" reading similar to a strong confirmation), although we did not come across any examples within the scope of our corpus. Again, we believe that these could be treated as clausal readings.
22. Whether this is desirable is less certain. G&S note that *echo* and *reference* reprise sentences are intonationally distinct, and this seems also true for sluices. It may be that although the content of both *can* always be expressed as clausal, there is good reason not to do so.
23. This runs contrary to our intuition which is that the gap form might have a constituent reading.

References

Allen, J. and Core, M. (1997). Draft of DAMSL: Dialog act markup in several layers.

Bohlin (Ljunglöf), P., Cooper, R., Engdahl, E., and Larsson, S. (1999). Information states and dialogue move engines. In Alexandersson, J., editor, *IJCAI-99 Workshop on Knowledge and Reasoning in Practical Dialogue Systems*.

Burnard, L. (2000). *Reference Guide for the British National Corpus (World Edition)*. Oxford University Computing Services.

Carletta, J. (1996). Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, 22(2):249–255.

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Fletcher, C. (1994). Levels of representation in memory for discourse. In Gernsbacher, M., editor, *Handbook of Psycholinguistics*. Academic Press.

Ginzburg, J. (1996). Interrogatives: Questions, facts and dialogue. In Lappin, S., editor, *The Handbook of Contemporary Semantic Theory*, pages 385–422. Blackwell.

Ginzburg, J. and Cooper, R. (2001). Resolving ellipsis in clarification. In *ACL/EACL01 Conference Proceedings*. Association for Computational Linguistics.

Ginzburg, J. and Cooper, R. (forthcoming). Clarification, ellipsis and utterance representation.

Ginzburg, J., Gregory, H., and Lappin, S. (2001a). SHARDS: Fragment resolution in dialogue. In Bunt, H., van der Sluis, I., and Thijsse, E., editors, *Proceedings of the Fourth International Workshop on Computational Semantics (IWCS-4)*, pages 156–172. ITK, Tilburg University, Tilburg.

Ginzburg, J. and Sag, I. (2000). *Interrogative Investigations: the Form, Meaning and Use of English Interrogatives*. Number 123 in CSLI Lecture Notes. CSLI Publications.

Ginzburg, J., Sag, I. A., and Purver, M. (2001b). Integrating conversational move types in the grammar of conversation. In Kühnlein, P., Rieser, H., and Zeevat, H., editors, *Proceedings of the Fifth Workshop on Formal Semantics and Pragmatics of Dialogue (BI-DIALOG 2001)*, pages 45–56.

Larsson, S., Ljunglöf, P., Cooper, R., Engdahl, E., and Ericsson, S. (2000). GoDiS - an accommodating dialogue system. In *Proceedings of ANLP/NAACL-2000 Workshop on Conversational Systems*.

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

Ross, J. R. (1969). Guess who? In Binnick, R. I., Davison, A., Green, G., and Morgan, J., editors, *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society*, pages 252–286. CLS, University of Chicago.

Sachs, J. D. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, 2:437–442.

Traum, D. (1994). *A Computational Theory of Grounding in Natural Language Conversation*. PhD thesis, University of Rochester.

van Dijk, T. A. and Kintsch, W. (1983). *Strategies of Discourse Comprehension*. Academic Press.