The Theory and Use of Clarification Requests in Dialogue

Matthew Purver

Submitted for the degree of Doctor of Philosophy
Department of Computer Science
King's College
University of London
August 26, 2004

Abstract

Clarification requests are an important, relatively common and yet under-studied dialogue device allowing a user to ask about some feature (e.g. the meaning or form) of an utterance, or part thereof. They can take many different forms (often highly elliptical) and can have many different meanings (requesting various types of information). This thesis combines empirical, theoretical and implementational work to provide a study of the various types of clarification request that exist, give a theoretical analysis thereof, and show how the results can be applied to add useful capabilities to a prototype computational dialogue system.

A series of empirical studies (corpus-based and experimental) are described which establish a taxonomy of the possible types of clarification request together with information about their meaning and usage, about the phrase types and conditions that trigger them and their particular forms and interpretations, and about the likely methods of responding to them.

A syntactic and semantic analysis using the HPSG framework is given which extends the work of (Ginzburg and Cooper, 2004) to cover the main classes of the above taxonomy, and to account for the clarificational potential of those word and phrase types which commonly cause clarification requests. This is shown to have interesting implications for the semantics of various lexical and phrasal types, in particular suggesting that noun phrases be given a simple witness-set based representation.

Finally, the theoretical analysis and empirical findings are applied within a HPSG grammar and a prototype text-based dialogue system, CLARIE. Implemented in Prolog using the TrindiKit, the system combines the information-state-based dialogue management of GoDiS (Larsson et al., 2000) and the HPSG-based ellipsis resolution of SHARDS (Ginzburg et al., 2001a) and adds the capability to interpret and respond to user clarification requests, and generate its own clarifications where necessary to deal with incomprehensible or contradictory input, resolve unknown or ambiguous reference, and learn out-of-vocabulary words.

Contents

List of Figures

List of Tables

Listings

Acknowledgements

First of all I owe a huge debt of gratitude to my supervisor, Jonathan Ginzburg, who has seen me through several years of vagueness and barking up wrong trees, helped stop my engineer's blunders into formal syntax and semantics from doing them too much damage, and generally been very helpful and supportive. He also got me involved with the ROSSINI project,[1] which proved to be the source not only of many of the results in here, but also of funding for three years via the Engineering and Physical Sciences Research Council. And not least, he provided the theoretical basis on which much of this thesis is built. Without that I wouldn't even have had anywhere to start.

I'm also grateful to the other project members, in particular Pat Healey but also the other members of his group at Queen Mary (James King, Greg Mills and Mike Thirlwell) who helped share the joy of running experiments. Thanks are also due to the others in the NLP group here at King's, with whom I've always enjoyed discussing my work, theirs and other random things – I'm thinking particularly of Shalom Lappin, Christian Ebert, Raquel Fernández and Leif Nielsen. The same goes for the Dynamic Syntax group, in particular Masayuki Otsuka, and even more so Ruth Kempson who has given me lots of thoughtful feedback and has always been remarkably generous in letting me concentrate on finishing this thesis when I really should have been working for her this year. Some of these random discussions have even ended up in Proper Collaborations, with publications and projects and stuff. Thanks Raquel, Masa and Ruth.

There are probably lots of other people who've helped me along the way (knowingly or not), and I'm not going to try to name them all, but I would like to thank Staffan Larsson and David Traum – their work has influenced a lot of my thinking, and they both always ask great questions at conferences. And a special mention goes to David Schlangen, who is not only able to discuss differing views without actually coming to blows, but can do it while finding the best bars and clubs in any given city.[2] Actually, I'd also like to thank Stephen Pulman, Ted Briscoe and Karen Spärck Jones, who got me interested in this field in the first place, back when I thought I wanted to study speech recognition.

Two more things. Firstly, a shout out to Anna in Kite: thank you for everything. Secondly, I'd like to dedicate this thesis to my mother. I know that's not a particularly original move, but I told her years ago that I wanted to do a PhD; she never got to see it, but I'm sure she would have been proud.

---

[1] `http://www.dcs.kcl.ac.uk/research/groups/gllc/projects/rossini.html`
[2] Except East Lansing, Michigan. But I don't think that's his fault.

Publications

Some of the work presented in this thesis has previously appeared in, or is due to appear in, various conference proceedings and journals. This is indicated in the relevant sections, and the details of the relevant publications are given here. In case 1 I was the sole contributor, and in cases 2–6 a major contributor, to the underlying research and to the paper as published. All other work presented here is my own, except where explicitly indicated otherwise.

1. **Matthew Purver.** Processing Unknown Words in a Dialogue System. In *Proceedings of the 3rd ACL SIGdial Workshop on Discourse and Dialogue*, pages 174–183, Philadelphia, July 2002.

2. **Matthew Purver and Jonathan Ginzburg.** Clarifying Noun Phrase Semantics in HPSG. In S. Müller, editor, *Proceedings of the 10th International Conference on Head-Driven Phrase Structure Grammar (HPSG-03)*, pages 338–358, East Lansing, July 2003.

3. **Matthew Purver and Jonathan Ginzburg.** Clarifying Noun Phrase Semantics. To appear in *Journal of Semantics*, 2004.

4. **Matthew Purver, Jonathan Ginzburg, and Patrick Healey.** On the Means for Clarification in Dialogue. In R. Smith and J. van Kuppevelt, editors, *Current and New Directions in Discourse & Dialogue*. Kluwer Academic Publishers, 2003. An earlier version also appeared in *Proceedings of the 2nd ACL SIGdial Workshop on Discourse and Dialogue*, pages 116–125, Aalborg, September 2001.

5. **Matthew Purver, Patrick Healey, James King, Jonathan Ginzburg, and Greg Mills.** Answering Clarification Questions. In *Proceedings of the 4th ACL SIGdial Workshop on Discourse and Dialogue*, pages 23–33, Sapporo, July 2003.

6. **Patrick Healey, Matthew Purver, James King, Jonathan Ginzburg, and Greg Mills.** Experimenting with Clarification in Dialogue. In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, Boston, August 2003.

Chapter 1

Introduction

1.1 Aims

This thesis has two main aims: to investigate and analyse the nature of clarification requests in dialogue in order to get a better understanding of how they arise, what forms they take, what they mean and how they are responded to; and to show how this understanding can lead to increased capabilities and more human-like behaviour in computational dialogue systems.

In consequence, the work described uses various different strategies: empirical, theoretical and implementational. The empirical work uses a corpus of dialogue, together with a new experimental setup, to gather evidence about naturally-occurring clarification. This evidence is then used to extend a basic theory of clarification, leading to a grammar fragment which covers and explains the results. This grammar and the empirical results then both feed into a prototype implementation, the CLARIE dialogue system.

In the process, an interesting and not entirely expected side-effect of the theoretical work has led to a secondary aim: to use clarificational phenomena to shed some light on the grammatical (and in particular semantic) properties of words and phrases in general. This may seem obvious (although it did not to me to begin with), but explaining the behaviour of clarification requests does not only require an examination of the requests themselves – we also have to examine and analyse the everyday words and phrases that are being clarified.

Two caveats are probably worth presenting to begin with. Firstly, the overall approach is what might be termed pragmatic (in a non-linguistic sense): rather than ensure that all possible phenomena are exhaustively categorised and analysed, the empirical work gives a sense of which phenomena are most common and most important, and this then feeds into both the theoretical and implementational parts, which attempt respectively to give analyses of these most important phenomena and show how such analyses can be used to build a prototype computational dialogue system. Secondly, the word 'prototype' in the preceding sentence is there for a reason.

1.2 Motivation

1.2.1 Linguistic

Clarification is a vital part of the human communicative process, and clarification requests (CRs) are resultingly common in human conversation. They can vary widely in surface form, as shown in example (1): from full explicit queries as in (a) or (b), to rather less explicit echoes such as (c), to more elliptical versions as in (d), to highly conventionalised fragments as in (e).

(1)

	A:	Did Bo leave?
	B:	(a) I'm sorry, what did you say?
		(b) Who do you mean 'Bo'?
		(c) Did BO leave?
		(d) BO?
		(e) Eh?

They can also vary widely in the clarificatory information that is being requested, and in how explicitly this is specified: in (a), what is being asked about is which words were uttered; in (b), the question seems to be more about semantic reference of one of those words. (c) and (d), on the other hand, seem more ambiguous: in different contexts and spoken with different intonations, they might be interpreted as asking whether B heard the word *Bo* correctly, or a question about the identity of this Bo. (e) seems even more difficult to pin down. The wide differences in what is being asked for become even clearer when considering what would constitute a felicitous answer: while a simple *"Yes"*, *"That's right"* or *"Uh-huh"* might be a perfectly suitable answer to (c) or (d) in many contexts, the others all expect something entirely different.

However, despite the fact that they have such different surface forms (which we shall refer to as CR *forms*) and seem to be asking different questions (which we shall refer to as CR *readings*), all of the above examples seem to have two things in common. Firstly, they all show that some sort of (partial) breakdown in communication has occurred: the participant initiating the CR exchange has some sort of problem with processing the previous utterance. Secondly, they are all in some sense *utterance-anaphoric* – they refer to the problematic utterance and query some feature of it.

This utterance-anaphoricity makes CRs an interesting area of linguistic study in their own right: their meaning seems to be of a different nature, or on a different level, to standard sentences. But it is the fact that in studying clarification we are studying *miscommunication* that makes the area even more worthwhile.

Much of linguistic theory makes a basic underlying assumption of perfect communication: formal semanticists, for example, are effectively interested in the content which would be associated with sentences by an ideal agent with perfect knowledge. While some of the phenomena above have been studied in some detail (echo questions like (c) being an exam-

ple), and developing suitable syntactic and semantic analyses is of course vital to a theoretical description, just giving such analyses often sheds little light on what causes such phenomena and how they fit into and affect the wider dialogue context.

Of course, people who study dialogue know that perfect communication is not the case – misunderstanding and miscommunication are common. If we wish to develop a theoretical understanding of linguistic processes which encompasses dialogue with all its attendant features, then, we must ensure that our theory can account for these breakdowns. Dialogue theorists have therefore often concentrated on how people manage the communicative process (e.g. Clark, 1996), and how this management or *grounding* process interacts with linguistic understanding (e.g. Poesio and Traum, 1997). This thesis takes one such approach as a starting point (Ginzburg and Cooper, 2004) to examine how a theoretical framework can be developed which not only allows us to describe clarification requests and assign meaning to them, but also enables us to model the processes which lead to clarification in the first place. From a theoretical point of view, then, it will address the following questions:

- What types of CRs exist, what do they mean, and can we give them a linguistic analysis?

- What types of word (or phrase, or sentence) can cause CRs, and what does this tell us about them?

- How can we model dialogue processes in such a way as to include miscommunication, clarification and its results?

1.2.2   Computational

Designers of computational dialogue systems have to take imperfect communication into account from the very beginning. Natural language being what it is, and computational natural language processing technology being what *it* is, a usual basic assumption is that the system has not perfectly understood what the user just said.

Many systems therefore use robust interpretation methods to attempt to get limited information from a user utterance – for example, spotting particular keywords which determine the utterance's intention, and throwing the rest away – or heavily restrict what the user can say at any particular point, thus making the task of interpretation much more predictable. However, the possibility of uninterpretability or uncertainty always exists, and even simple slot-filling systems will always have the ability to indicate this in some general way: *"I don't understand"* or *"Please repeat that."*

As systems have started dealing with more complex tasks and domains, and thus requiring more detailed and complete representations of meaning to be extracted from what the user says, the scope for various forms of misunderstanding increases: words which cannot be parsed by the grammar, ambiguous words and sentences, problems with reference resolution.

Further clarificational abilities have therefore started to be added: some systems can, say, report which unknown word is causing it a problem (Hockey et al., 2002), or ask about the intended reference of a pronoun (Traum et al., 2003).

In other words, the need for systems to be able to clarify user utterances is well established. However, the methods used to generate CRs and then to deal with the ensuing clarificational dialogue are usually somewhat *ad hoc*, with system prompts designed as seem right at the time, and subsequent dialogue dealt with by a separate module specifically designed for, say, unknown word clarification and acquisition (Knight, 1996). There is therefore a need firstly for a study of CRs in dialogue to determine which types of CRs are good at clarifying and eliciting which kind of information, and secondly for a framework which will allow all types of clarification to be dealt with in a uniform way within the standard dialogue processes. The questions that need to be answered are as follows:

- How should systems generate CRs such that they are suitably interpreted and responded to by the user?

- How (and when) should we expect users to respond to these system CRs?

- How can CRs and their responses fit smoothly into the general dialogue, so that it can continue once the required information is provided?

In contrast, it is not usual for computer dialogue systems to be able to process CRs produced by the user. Designers (very sensibly) attempt to avoid having to deal with user CRs by making system prompts as clear and informative as possible, and if possible by training users to use the system in a particular way and to know what to expect. However, as systems start to get a wider audience, and start to deal with wider domains and more complex tasks, and indeed as they become more human-like and begin to be treated more like humans by users (see 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. 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 user:

(2)
System:	Would you like to travel 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 to the user: response (d), which incorrectly interprets it as an *answer* to the system's question, would not be acceptable under any circumstances. However, without

the capability to recognise 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).

Which of (a)–(c) is actually helpful will depend on the reading intended: the question that the CR was intended to ask. Dealing with user CRs properly, then, will require not only the ability to recognise an utterance as a CR, but some ideas about which forms are associated with which readings in which contexts. In other words, there are two further interesting questions to ask:

- How can a system recognise and correctly interpret a user CR?

- How (and when) should a system answer a user CR?

Another Caveat These, then, are some of the questions that this thesis *will* address. At this point it is worth mentioning one aspect of clarification that will *not* be taken up here: that of levels of confidence. Given the uncertain nature of speech recognition and the continuous nature of the probability estimates it is based on, it is not trivial for a dialogue system to decide when recognition confidence is low enough to require clarification, and when it is not. In other words, another question that needs to be answered in practical applications is:

- When should a system generate a CR at all, rather than assuming all is well?

In practice, systems usually decide whether to clarify user utterances based on their speech recognition confidence scores. Above a certain upper threshold, utterances can be accepted as they are; below a lower threshold, a CR should be generated (e.g. a general *"What did you say?"*). In between these levels, the system may choose to check that its interpretation is correct, for example by *explicitly* asking for confirmation (*"I think you said you wanted to go to Paris. Is that correct?"*) or perhaps, if slightly more confident, including confirmation *implicitly* in its next utterance (*"So, what time do you want to go to Paris?"*). Such techniques have been examined and used by, amongst others, San-Segundo et al. (2001) and Larsson (2002), and have more recently been extended to combine speech recognition and pragmatic plausibilities by Gabsdil and Bos (2003) and Schlangen (2004).

In the work that follows, though, this question will not be addressed except in a very simplistic way. The assumption will be that a deterministic decision can always be made – that some utterance (or one of its constituent parts) either requires clarification or it does not. Of more interest will then be the questions of what exactly can be clarified and how, as set out above. In real applications, the familiar probability thresholding techniques will also be required, but this seems entirely compatible with the approach which will be set out below.

1.3   Structure

The basic structure of the thesis is first to examine empirical evidence, then to show how a theoretical framework can be extended accordingly, and finally to show how the theoretical model can be implemented within a computational grammar and dialogue system which reflect the empirical findings and provides some of the desired clarificational capability. The chapters are arranged as follows:

- **Chapter 2 (Background)** introduces some notation and gives a review of some relevant linguistic and computational work. In particular, Ginzburg and Cooper (2004)'s syntactic, semantic and contextual analysis of reprise sentences and fragments is introduced, which is used as the basis of the theoretical and grammatical analyses proposed throughout. An introduction to Larsson et al. (2000)'s GoDiS dialogue system is also given, which is used as the basis of the implemented dialogue system, together with the TrindiKit toolkit which it uses.

- **Chapter 3 (Empirical Observations)** presents several corpus-based and experimental studies into the empirical nature of clarificational dialogue. An initial corpus study provides a taxonomy of CR types together with statistical information about their meaning and usage. Further corpus studies then investigate the nature of the sources of CRs and of responses to CRs. Finally, an experimental study provides some more fine-grained results concerning the relation between CR, source and response type for a certain common class of CR.

- **Chapter 4 (Implications for Semantics)** extends the theoretical work of Ginzburg and Cooper to account for the clarificational potential of the word and phrase types found to be the most common sources of CRs. Significantly, this results in a syntactic and semantic analysis of noun phrases which departs from traditional higher-order and generalised quantifier views, using instead a flat witness-set based representation.

- **Chapter 5 (A Grammar for Clarification)** shows how this approach can be implemented within a HPSG grammar to give suitable analyses for the various types of CR found in chapter 3, including a treatment of elliptical fragments which is integrated into a general view of contextual dependence.

- **Chapter 6 (The CLARIE System)** describes how the grammar and empirical findings can be implemented and used within an information-state-based dialogue system. The resulting prototype system, CLARIE, has the capability to interpret and respond to user-initiated clarificational dialogue, and to initiate such dialogue itself when faced with problematic input, allowing it to interactively resolve reference and ambiguity, and learn out-of-vocabulary words.

- **Chapter 7 (Conclusions and Future Directions)** then summarises the main findings of the previous chapters and draws some overall conclusions about clarification and linguistic representation with particular regard to dialogue systems. Some future research directions are then briefly discussed.

Chapter 2

Background

2.1 Overview

This chapter first introduces the notation used for the repesentation of questions, and for linguistic analysis in general, in section 2.2. It then describes some relevant previous work on clarification in dialogue, firstly from what might broadly be termed a linguistic viewpoint in section 2.3, and then from a more computational and implementational perspective in section 2.4.[1] Section 2.5 then gives some background information on two implemented systems on which the CLARIE system builds, GoDiS and SHARDS.

2.2 HPSG Notation

The linguistic analysis given throughout this thesis, derived as it is from (Ginzburg and Cooper, 2004), is couched in HPSG (Pollard and Sag, 1994), and in particular the version described in (Ginzburg and Sag, 2000). Particular analyses will therefore be shown as feature structures, or attribute-value matrices (AVMs). Space prohibits a full explanation of HPSG AVM notation here (the reader is referred to (Ginzburg and Sag, 2000)); however, in an attempt both to save space and to make examples more readable for those not familiar with HPSG, abbreviations will be used wherever possible for the most common types of semantic object that will be discussed. These are shown in table 2.1.

For *parameters* (the semantic content of noun phrases – and, as will be proposed later, of some other phrase types) and for *propositions*, the abbreviated notation should be relatively self-explanatory. However, given the particular importance of *questions*, some more detail may be warranted. In the framework of (Ginzburg and Sag, 2000), questions are regarded as abstracts, with a set of queried parameters PARAMS simultaneously abstracted from

---

[1]Much of the work described here could of course be classified as belonging to both sections, making computational advances and contributing to linguistic theory. The division is somewhat arbitrary and more due to how they relate to this thesis – apologies for any misrepresentation.

AVM	Abbreviation

$$\begin{bmatrix} parameter \\ \text{INDEX} \quad x \\ \text{RESTR} \quad \left\{ \begin{bmatrix} \text{INSTANCE} & x \\ \text{PROPERTY} & P \end{bmatrix} \right\} \end{bmatrix}$$

$$x:property(x,P)$$

The semantic content of a referential noun phrase: an object of type *parameter* consisting of an index x with a restriction expressed as the fact that a particular property P holds of it. For example, the proper name *John* will be given the content $j : name(j, john)$. The treatment of quantification will be left aside for now, and given later in chapter 4.

$$\begin{bmatrix} proposition \\ \\ \text{SOA} \mid \text{NUCLEUS} \quad \begin{bmatrix} verb\_rel \\ \text{ROLE\_1} & x \\ \text{ROLE\_2} & y \end{bmatrix} \end{bmatrix}$$

$$verb(x,y)$$

A proposition containing a verbal relation $verb$ which holds between two indices x and y. For example, the proposition expressed by the sentence *"John likes Mary"* will be given as $like(j, m)$, where the contents associated with *John* and *Mary* are $j : name(j, john)$ and $m : name(m, mary)$ respectively. Again, details of quantification will be given in chapter 4.

$$\begin{bmatrix} question \\ \text{PARAMS} \quad \{\} \\ \text{PROP} \quad verb(x,y) \end{bmatrix}$$

$$?.verb(x,y)$$
$$\text{or}$$
$$?\{\}.verb(x,y)$$

A polar (yes/no) question, formed by the abstraction of an empty PARAMS set from the propositional body PROP. Continuing the example above, the question expressed by the sentence *"Does John like Mary?"* will be written as $?.like(j, m)$.

$$\begin{bmatrix} question \\ \text{PARAMS} \quad \{x:property(x,P)\} \\ \text{PROP} \quad verb(x,y) \end{bmatrix}$$

$$?x.verb(x,y)$$
$$\text{or}$$
$$?\{x:property(x,P)\}.verb(x,y)$$

A *wh*-question, formed by the abstraction of a non-empty PARAMS set containing the queried parameter. The question expressed by the sentence *"Who likes Mary?"* will be written as $?x.like(x, m)$, or, including the restriction that the queried parameter associated with *Who* must refer to a person, $?\{x:person(x)\}.like(x, m)$.

Table 2.1: HPSG AVM Abbreviations

a propositional body PROP. Formally, these simultaneous abstracts are objects of a specific type *question* in a situation-semantic universe (see Ginzburg and Sag, 2000, for more details), but they can be thought of as roughly equivalent to λ-abstracts: thinking of the question $?\{x\}.verb(x, y)$ as the λ-abstract $\lambda x.verb(x, y)$ will not be too far off the mark. Examples like this which have a non-empty abstracted set correspond to *wh*-questions; polar yes/no questions simply have an empty abstracted set.

2.3 Linguistic Approaches

In this section, theoretical work in the linguistic and conversation-analytic traditions is introduced, with particular reference to the recent work of Jonathan Ginzburg and his collaborators, which is taken as the departure point for the formal framework used in this thesis.

2.3.1 Grounding and Feedback

Clarificational dialogue has been recognised as an important feature of human dialogue for some time, with two approaches in particular bringing it to the fore. Firstly, work in the conversation-analytic tradition such as that of Schegloff (1987) and Sacks (1992) introduced the notion of the *back-channel* – by treating dialogue as existing on several different levels, where each is separate from (but possibly concerning) the previous one, the initiation of clarificational or repair dialogue could be seen as the move to the next level, returning when the repair discussion is complete. This method allows dialogues to be analysed in terms of these levels and the importance of back-channels and repairs to be seen, but gives no systematic way of analysing the particular contribution of a turn (say, the meaning of or question posed by a particular CR).

Secondly, the work of Clark (1992, 1996) and Allwood (2000) brought out a focus on the basic need to establish and maintain successful communication. Both see one of the primary tasks of a conversational participant (CP) as monitoring the conversation for evidence that the communication is successful and that the immediately preceding utterance has been understood or *grounded* at various levels. Exact definitions of the levels differ; (Allwood, 2000) gives a version as follows:

1. contact – the will or ability to continue the interaction;

2. perception – the will or ability to perceive the message;

3. understanding – the will or ability to understand the message;

4. attitudinal reactions – the will or ability to react or respond to the message, including acceptance or rejection.

Feedback (or to use Allwood's term, interactive communication management) can then be seen as exchange of information about any of these requirements (see Allwood et al., 1992). This definition of feedback includes acknowledgements, gestures and so on, but also encompasses clarification: CRs perform negative feedback actions at one or more levels, indicating lack of ability to perceive (or understand etc.) the message. The existence of these different levels of grounding can then give some insight into what the various meanings of CRs might be: some might give feedback about the perception level (e.g. *"What did you say? / Pardon?"*; some at the understanding level (e.g. *"What? / What do you mean?"*); and some at the level of acceptance/rejection (e.g. *"Really? / Bo?"*).

These insights therefore take us a step further in that they can give some idea of the various possible functions or meanings of CRs, and correspondingly some idea of what causes of CRs might be, but still do not afford any way of extracting specific meanings from individual utterances.[2] For this we must look to linguistic and grammatical theory.

2.3.2 Metalinguistics and Metarepresentation

Interest in the syntax and semantics of clarification has been limited – linguists have often treated CRs as *metalinguistic* and therefore outside the scope of normal grammar. Those who do approach the subject have mainly focussed on reprise sentences or *echo questions*, as they appear to have idiosyncratic syntactic and semantic characteristics: while expressing an interrogative content, they have seemingly declarative structure, requiring neither *wh*-fronting nor subject-auxiliary inversion (example (3)); they can focus on units below the word level (examples (4) and (5)); and as with all clarification, their content appears to have some metalinguistic nature, asking about some property of a previous utterance.

(3)[3] | A: I've bought you an aeroplane.
         | B: You've bought me an AEROPLANE?
         | B':   You've bought me a WHAT?

(4)[4] | A: Have you seen my agapanthus?
         | B: Have I seen your aga-WHAT?

(5)[5] | A: I've been reading a bit recently about [*auditory disturbance in the room*]jacency.
         | B: Sorry, you've been reading about WHAT-jacency?

This led to early proposals such as that of Janda (1985) which sees echoes as fundamen-

---

[2]More recently, Schlangen (2004) extends this approach to a more fine-grained model of levels of communication and specifically links these levels to causes of CRs, but again provides no connection between these causes and the surface form or meaning of individual CRs.

[3]From (Blakemore, 1994).

[4]From (Blakemore, 1994), p.202.

[5](Ginzburg and Sag, 2000)'s example (5), p.257.

tally different from standard questions: as questions about surface strings, with *wh*-words substituted for arbitrary sequences of syllables in the previous utterance. String based approaches like this quickly run into trouble, however: as the examples above show, faithful copying of the string is not only not required (example (5) misses out *"a bit"* in the echo) but often prohibited – all the examples above feature substitution of indexicals (*I* with *You*, *my* with *your* and so on; and semantically co-referential but phonologically distinct expressions can generally be substituted:

(6)[6]
A:	Rusty chewed the antique chair you lent us.
B:	Your dog chewed WHAT?

More semantically-based approaches are those of Blakemore (1994), Noh (1998, 2001) and most recently Iwata (2003) who, broadly speaking, give a *metarepresentational* account of echo questions couched in Relevance Theory (Sperber and Wilson, 1986). On this account, an echo question metarepresents the previous utterance by exploiting similarities in surface form (what Iwata calls a metalinguistic use) or in content (a metaconceptual use). This is then understood as asking a particular question about a feature of the previous utterance (for Noh, its content including its illocutionary force, while Iwata includes other possibilities including its surface form). However, this question must be identified from the metarepresentation via a process of pragmatic inference, and so this approach sheds little light on what the actual possible meanings of CRs are, and is also not ideal for a computational implementation.

2.3.3   Focus Semantics

Artstein (2002), building on work of Hockey (1994), gives a more formal semantic analysis of echo questions based on focus semantics and the alternative set approach to the semantics of questions (Groenendijk and Stokhof, 1984). An non-*wh* echo question such as the response marked B in example (3) is taken to denote the set of all propositions derived by substituting alternatives to the denotation of the focussed constituent (in this case, *AEROPLANE*); similarly, a *wh*-echo such as the response marked B' denotes the set of all propositions derived by substituting alternatives for the *wh*-phrase. Semantically, then, both of these echo questions are indistinguishable from each other, and from the direct non-echo question *"What have you bought me?"*. He sees the difference between the echo and direct versions, and the difference between the *wh*- and non-*wh* versions, as being purely pragmatic: while a direct question asks for a true proposition, an echo asks for the proposition that was intended by the original speaker; and while a *wh*-echo merely asks a question, a non-*wh* echo also offers a proposition, thus expressing surprise or some other attitude towards it. In addition, and importantly, the semantic meaning of the echo must be derived pragmatically, by Gricean inference, from the standard propositional semantics given by its seemingly declarative syntax. Again, if we are

---

[6](Artstein, 2002)'s example (41), p. 104

looking for a method of analysis that lends itself to computational implementation, this heavy reliance on pragmatic inference is a significant disadvantage.

However, implementation aside, Artstein's analysis does have a potentially useful feature: it allows focus below the word level, and therefore allows a formal analysis of examples such as examples (4) and (5), with a full semantics (rather than a (Janda, 1985)-like question about the surface string).

2.3.4   Ginzburg and Sag

Ginzburg and Sag (2000) (hereafter G&S) take a different approach to reprises, treating them not as metalinguistic but rather as syntactically standard sentences of their grammar, and giving them a grammatically-assigned semantic analysis rather than using reasoning or inferential processes. Their primary arguments for this stance are firstly that the non-fronted syntactic form of reprise interrogatives is not limited to reprises, but can be used for perfectly standard *direct* questions (see example (7)); and secondly that their semantic content can be adequately expressed by standard interrogatives, as long as they can be taken as referring to the content of the original utterance being reprised (see examples (8) and (9) and their accompanying suggested paraphrases). Their analysis therefore acknowledges the utterance-referring nature of reprises while not classing them as metalinguistic or metarepresentational.

(7)[7]   A: I'm going to send the sourdough bread to the Southern Bakery, and the croissants to Barringers.
       B: I see, and the bagels you're going to send WHERE?

(8)[8]   A: Chris is annoyed with Jan.
       B: Chris is annoyed with WHO(M)?
       ↝   *"Who did you assert/say that Chris is annoyed with?"*

(9)[9]   A: Merle attacked Brendan yesterday.
       B: Merle attacked Brendan yesterday? / attacked Brendan?
       ↝   *"Did you assert/say that Merle attacked Brendan yesterday?"*
       B':   Merle? / Attacked? / Brendan? / Yesterday?
       ↝   *"Is it **Merle** that you said attacked Brendan yesterday?"*
       ↝   *"Is it **attacked** that you said Merle did to Brendan yesterday?"*

They offer a syntactic and semantic analysis, using HPSG (Pollard and Sag, 1994): 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 shown in the paraphrases of examples (8) and (9), this content includes the illocutionary force of that previous utterance. Their grammar therefore includes that illocution-

---

[7]G&S's example (65), p.280
[8]G&S's example (16), p.260
[9]G&S's example (10), p.259

ary force as part of the semantic representation of all sentences. They posit a semantic type *illoc(utionary)-rel(ation)*, which has four subtypes *assert, ask, order* and *exclaim* as shown in the hierarchy of figure 2.1:

illoc-rel

assert-rel   *ask-rel*   *order-rel*   *exclaim-rel*
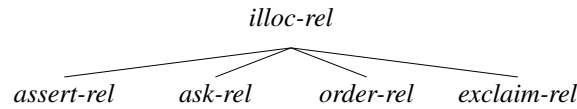
Figure 2.1: Basic Conversational Move Type Hierarchy

These relations are three-place relations between a speaker, an addressee and a message (in the case of asserting, the proposition being asserted; in the case of asking, the question being asked, etc.). Their top-level clause type *root-cl(ause)* is then specified to have a semantic content containing such an *illoc-rel* relation, with the type and message content determined by the syntactic and semantic form of the sentence. This means that for an assertion such as A's in example (8), the sentence denotes the proposition that A has asserted something to B, where that something is the proposition that Chris is annoyed with Jan:

(10) $assert(A, B, annoyed(C, J))$

The representation of the sentence therefore takes the form of AVM (11), or in abbreviated form in AVM (12):

$$(11) \begin{bmatrix} root\text{-}cl \\ \text{PHON} \quad \langle \text{chris, is, annoyed, with, jan} \rangle \\ \text{CONTENT} \begin{bmatrix} proposition \\ \text{SOA} \mid \text{NUCL} \begin{bmatrix} assert\text{-}rel \\ \text{UTTERER} \quad A \\ \text{ADDRESSEE} \quad B \\ \text{MSG-ARG} \begin{bmatrix} proposition \\ \text{SOA} \mid \text{NUCL} \begin{bmatrix} annoyed(C, J) \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

$$(12) \begin{bmatrix} root\text{-}cl \\ \text{PHON} \quad \langle \text{chris, is, annoyed, with, jan} \rangle \\ \text{CONTENT} \begin{bmatrix} assert(A, B, \begin{bmatrix} annoyed(C, J) \end{bmatrix}) \end{bmatrix} \end{bmatrix}$$
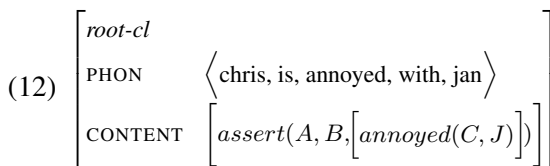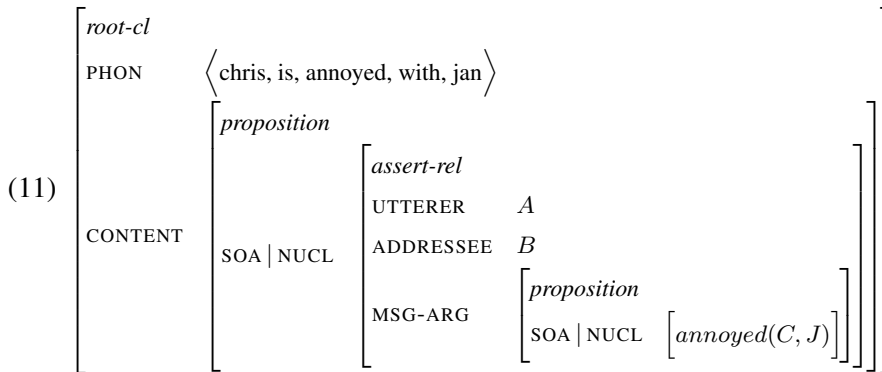
Reprises can then be specified as having a content which is a question about the perceived content of a previous utterance (provided that this is compatible with the content directly assigned by their syntactic daughters) and this will give rise to a suitable reading including

the previous utterance's illocutionary force.[10]

$$(13) \quad \begin{bmatrix} \textit{reprise-int} \\ \text{CONTENT} \quad \begin{bmatrix} \textit{question} \\ \text{PROP} \quad \boxed{1}\begin{bmatrix} \text{SOA} \mid \text{NUCL} \mid \text{MSG-ARG} & \boxed{2} \end{bmatrix} \end{bmatrix} \\ \text{HEAD-DTR} \mid \text{CONTENT} \quad \boxed{2} \\ \text{BACKGRND} \quad \left\{ \textit{previous-utt}(\boxed{0}), \textit{perceived-content}(\boxed{0},\boxed{1}) \right\} \end{bmatrix}$$

For example (8), where the previous utterance is A's initial utterance which has the representation of AVM (12), this will ensure that the question concerns the previous assertion $assert(A, B, annoyed(C, J))$:

$$(14) \quad \begin{bmatrix} \textit{reprise-int} \\ \text{PHON} \quad \left\langle \text{chris, is, annoyed, with, who} \right\rangle \\ \text{CONTENT} \quad \begin{bmatrix} \textit{question} \\ \text{PARAMS} \quad \left\{ J \right\} \\ \text{PROP} \quad \boxed{1}\begin{bmatrix} assert(A, B, \boxed{2}\begin{bmatrix} annoyed(C, J) \end{bmatrix}) \end{bmatrix} \end{bmatrix} \\ \text{HEAD-DTR} \mid \text{CONTENT} \quad \boxed{2}\begin{bmatrix} annoyed(C, J) \end{bmatrix} \\ \text{BACKGRND} \quad \left\{ \textit{previous-utt}(\boxed{0}), \textit{perceived-content}(\boxed{0},\boxed{1}) \right\} \end{bmatrix}$$

In this case, the syntactic and semantic properties of the *wh*-word *who* ensure that its associated parameter J becomes a member of the abstracted (queried) PARAMS set, making the content a question paraphraseable *"Who$_J$ are you asserting that Chris is annoyed with J?"* as desired. Of course, this content (a question) then becomes embedded at *root-clause* level within another layer of illocutionary force, giving as overall content the proposition that B has *asked* A this question.

Crucially, this analysis starts to give what we are looking for – a syntactic and semantic analysis which is defined within a grammar which lends itself to computational implementation, and which does not require heavyweight inference. Of course, it only covers one type of clarification request, the reprise sentence (with or without a *wh*-phrase), and does not say much about possible causes of clarification (although it does require reprises to be associated with suitable previous utterances).

2.3.5   Ginzburg and Cooper

G&S extend the analysis to two elliptical forms, which they term *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;

---

[10]In G&S's analysis, the conditions on the previous utterance are expressed through a BACKGRND feature, a set of facts which must hold in context. This particular feature will not be used hereafter, with its function being fulfilled by C-PARAMS – see section 2.3.5.

as shown in example (15), this interpretation could be paraphrased either as the full (non-elliptical) reprise sentence, or as an equivalent standard (non-reprise) interrogative expressing the same content:

(15)
A: Did Bo leave?
B: **Who?**
$\leadsto$ (non-elliptical: *"Did **who** leave?"*)
$\leadsto$ (non-reprise: *"Who$_i$ are you asking me whether i left?"*)

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

(16)
A: Did Bo leave?
B: **Bo?**
$\leadsto$ (non-elliptical: *"Did **Bo** leave?"*)
$\leadsto$ (non-reprise: *"Is it **Bo**$_i$ that you're asking me whether i left?"*)

Their approach to ellipsis resolution is based on the idea of questions under discussion (QUDs) and follows that of SHARDS (Ginzburg et al., 2001a). More details of this approach are given below in section 2.5.2, but in brief, elliptical fragments are defined in the grammar (via particular construction types) to take as their content a proposition or question derived not only from their constituent word(s) but from a maximally available contextual question, the maximal QUD or MAX-QUD. Asking an explicit question (e.g. *"Who left?"*) will raise this question (i.e. $?x.leave(x)$) as the maximal QUD. This then allows a subsequent fragment to form its own propositional content from this question, while filling in the role played by the queried parameter; a bare answer *"Bo"* would be assigned the content $leave(b)$, or *"Bo left"*.

Resolution of the two elliptical reprise forms can now be achieved by allowing a conversational participant to coerce a clarification question onto the list of QUDs in the current dialogue context, without it being explicitly asked. As long as the possible coerced QUDs are suitable, this allows ellipsis resolution in the same manner as outlined above to give essentially the same reading as the non-elliptical reprise forms.

Ginzburg and Cooper (2001, 2004) (hereafter G&C) then take this extension further, giving a more detailed analysis for the bare fragment form (therein described as *clarification ellipsis*, CE), supplying more details of the possible QUD coercion mechanisms including a further semantic reading, and also specifying a model of the contextual dependence of utterances which allows an analysis of how clarification questions arise.

Contextual Parameters

Standard versions of HPSG directly encode idealised semantic content (that which a speaker would be expected to associate with a sign) within the value for the CONTENT feature. Instead, G&C propose a representation which expresses contextual dependence, one which encodes

meaning rather than *content*: a function from context to fully specified content. Contextually dependent parameters such as speaker, hearer, utterance time and (crucially to their analysis of clarification) the reference of proper names are abstracted to a set which is the value of a new C-PARAMS feature, as shown in AVM (17) for A's original utterance in example (16):

$$(17) \quad \begin{bmatrix} \text{PHON} & \left\langle \text{did, bo, leave} \right\rangle \\ \text{CONTENT} & \left[ask(a, b, ?.leave(x)) \right] \\ \text{C-PARAMS} & \left\{ [a\!:\!speaker(a)], [b\!:\!addressee(b)], [x\!:\!name(x, Bo)] \right\} \end{bmatrix}$$

Such representations of meaning can be viewed as λ-abstracts, with the members of C-PARAMS simultaneously abstracted over the standard value of CONTENT. More specifically, they are interpreted as *simultaneous abstracts with restriction* as shown in (18): $\{ABS\}$ 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 further formal details, see (Ginzburg and Sag, 2000).

$$(18) \quad \lambda\big\{ABS\big\}[RESTR].BODY$$

AVM (17) can therefore be rewritten as in (19), or, simplifying even further by omitting the parameters associated with speaker and addressee, as in (20). Wherever possible (mainly in chapter 4), these equivalent λ-abstract expressions will be included for readability's sake.

$$(19) \quad \lambda\big\{a, b, x\big\}[speaker(a), addressee(b), name(x, Bo)].ask(a, b, ?.leave(x))$$

$$(20) \quad \lambda\big\{x\big\}[name(x, Bo)].ask(a, b, ?.leave(x))$$

These utterance-level representations are built up compositionally[11] by the grammar. Lexical items such as proper names are defined to introduce abstracted parameters in C-PARAMS – the word *Bo* is given the representation below:

$$(21) \quad \begin{bmatrix} \text{PHON} & \left\langle \text{bo} \right\rangle \\ \text{CONTENT} & \boxed{1}\,x\!:\!name(x, Bo) \\ \text{C-PARAMS} & \left\{ \boxed{1} \right\} \end{bmatrix}$$

These parameters are then inherited via a C-PARAMS *amalgamation principle*: the value of C-PARAMS for lexical heads is defined to be the set union of the values of its syntactic sisters, and this is inherited up via heads to the sentence level. This gives the correctly contextually dependent meaning for the whole utterance, as shown in (22) for the sentence of

---

[11]The grammar uses various *constructions* which define how meaning is built up from constituent parts: this may not be consistent with some strict definitions of the principle of compositionality, but is compositional according to definitions such as that of (Pelletier, 2003) – the grammar gives a principled procedure for establishing utterance meanings given lexical items and their syntactic mode of combination.

example (16).

(22)

$$
\begin{bmatrix}
\text{PHON} & \langle \text{did, Bo, leave} \rangle \\
\text{CONTENT} & \left[ask(?.\boxed{1}) \right] \\
\text{HEAD-DTR} & \boxed{3} \\
\text{C-PARAMS} & \boxed{P} = \left\{ \boxed{2} \right\}
\end{bmatrix}
$$

$$
\boxed{3}\begin{bmatrix}
\text{PHON} & \langle \text{did} \rangle \\
\text{CONTENT} & \boxed{1} \\
\text{C-PARAMS} & \boxed{P} = \boxed{P_1} \cup \boxed{P_2}
\end{bmatrix}
\quad
\begin{bmatrix}
\text{PHON} & \langle \text{Bo} \rangle \\
\text{CONTENT} & \boxed{2}\left[x : name(x, Bo) \right] \\
\text{C-PARAMS} & \boxed{P_1}\left\{ \boxed{2} \right\}
\end{bmatrix}
\quad
\begin{bmatrix}
\text{PHON} & \langle \text{leave} \rangle \\
\text{CONTENT} & \boxed{1}\left[leave(x) \right] \\
\text{C-PARAMS} & \boxed{P_2}\{\}
\end{bmatrix}
$$
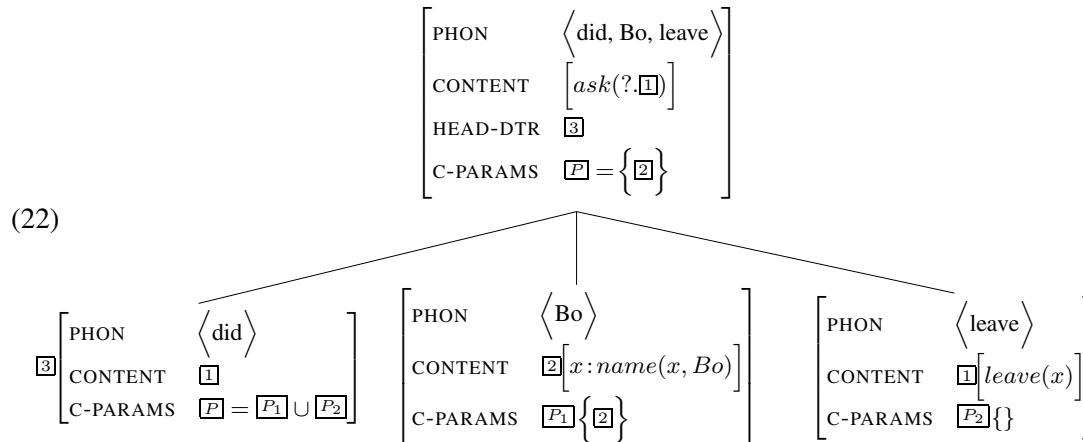
Grounding and Reprises

The grounding process for an addressee can now be modelled as an application of this mean-
ing abstract to the context, establishing the referents of the abstracted parameters such that
their given restrictions are satisfied, and resulting in the full fixed semantic content. It is
failure do this for a particular parameter that results in the formation of a clarification ques-
tion, with the purpose of querying the sub-utterance which contributed that parameter. Failure
may be due to, say, the lack of an available referent in context (e.g. no known person named
Bo), the lack of a unique most salient referent (e.g. two equally salient people named *Bo*), or
an available referent which is problematic in some way (e.g. leading to inconsistency in the
resulting content). This model, then, explains how CRs arise: they are triggered by contex-
tually dependent parameters which have been abstracted from content but which cannot be
instantiated in the current context.

   The resulting clarification question can take many forms: direct non-reprise questions
(*"Who do you mean by 'Bo'?"*), reprise echo questions (*"Did BO leave?"*), and elliptical
reprise sluices (*"WHO?"*) and bare fragments (*"BO?"*) are some of the possibilities which
they give analyses for.

Contextual Coercion and Resolution

G&C give a QUD-based analysis of how the content of a CR is derived in context: rather than
relying on general pragmatic inference, they take a conversational participant's basic dialogue
competence to include certain specific contextual update tools or *coercion operations*, which
take the utterance being clarified as their input and produce a partially updated context where
this utterance is salient and the maximal QUD is a suitable clarification question. They define

two such possible operations – here we describe them and show how they lead to suitable contents being assigned to the most complex example, the bare fragment *"BO?"*.

Clausal Readings   In the case when a hearer finds a problematic value for a contextual parameter, the question that arises is a *clausal* question, a *polar* (yes/no) question about the parameter's intended referent, corresponding to the first of the paraphrases given in example (16) above or to that given in example (23):

(23)
A:	Did Bo leave?
B:	**Bo? / Bo Smith?**
A:	That's right.
B:	Yes, half an hour ago.

$\leadsto$ *"Is it **Bo**$_x$ / **Bo Smith**$_x$ that you are asking whether x left?"*

As shown, reprises with clausal readings can repeat the original phrase verbatim (*"Bo?"*) or can use another apparently co-referring phrase (*"Bo Smith?"*). We will call verbatim repeats *direct echoes*.

The coercion operation for these readings produces an updated context where the new maximal QUD is the question formed by abstracting the problematic parameter from the original intended content, and the new salient utterance is the (sub-)constituent associated with that problematic parameter, as shown in AVM (24):

(24)
$$
\begin{bmatrix}
\text{C-PARAMS} & \{\dots, \boxed{1}, \dots\} \\
\text{CONSTITS} & \{\dots, \boxed{2}[\text{CONTENT} \quad \boxed{1}], \dots\} \\
\text{CONTENT} & \boxed{3}
\end{bmatrix}
\qquad \text{(original utterance)}
$$
$$
\Rightarrow
$$
$$
\begin{bmatrix}
\text{CONTEXT} & \begin{bmatrix} \text{SAL-UTT} & \boxed{2} \\ \text{MAX-QUD} & ?\boxed{1}.\boxed{3} \end{bmatrix}
\end{bmatrix}
\qquad \text{(partial reprise context description)}
$$

In the case of example (23), the problematic parameter is the content of the sub-constituent *Bo*, and the resulting QUD becomes $?\{x:name(x,Bo)\}.ask(a,b,?.leave(x))$, paraphrasable as *"For which Bo$_x$ are you asking whether x left?"*). The salient utterance SAL-UTT is the original sub-utterance *Bo*:

(25)
$$
\begin{bmatrix}
\text{C-PARAMS} & \{\dots, \boxed{1}, \dots\} \\
\text{CONSTITS} & \left\{\dots, \boxed{2}\begin{bmatrix} \text{PHON} & \langle \text{bo} \rangle \\ \text{CONTENT} & \boxed{1}[x:name(x,Bo)] \end{bmatrix}, \dots\right\} \\
\text{CONTENT} & \boxed{3}ask(a,b,?.leave(x))
\end{bmatrix}
\qquad \text{(original utterance)}
$$
$$
\Rightarrow
$$
$$
\begin{bmatrix}
\text{CONTEXT} & \begin{bmatrix} \text{SAL-UTT} & \boxed{2} \\ \text{MAX-QUD} & ?\boxed{1}.\boxed{3} \end{bmatrix}
\end{bmatrix}
\qquad \text{(partial reprise context description)}
$$

Their grammar defines elliptical bare fragments as having a content which is determined by the contextual MAX-QUD and SAL-UTT features. This is described in detail in section 2.5.2 below, but AVM (26) gives a sketch for the fragment *"Bo?"* – its content is a polar interrogative formed from some MAX-QUD wh-question in which *Bo* is the queried parameter:[12]

$$(26) \quad \begin{bmatrix} \text{CONTENT} & ?.\boxed{3} \\ \text{HEAD-DTR}\,|\,\text{CONTENT} & \begin{bmatrix} x\!:\!name(x, Bo) \end{bmatrix} \\ \text{CONTEXT} & \begin{bmatrix} \text{SAL-UTT}\,|\,\text{CONTENT} & \begin{bmatrix} x \end{bmatrix} \\ \text{MAX-QUD} & ?x.\boxed{3} \end{bmatrix} \end{bmatrix}$$

As shown in AVM (27), the new context produced by the coercion operation of AVM (25) now directly causes this fragment to be resolved as having the content $?.ask(a, b, ?.leave(x))$ (paraphrasable as in example (23) above). A reprise sluice *"Who?"* would be resolved similarly, but would also contribute a queried *wh*-parameter, thus making its content identical to the new maximal QUD.

$$(27) \quad \begin{bmatrix} \text{CONTENT} & ?.\boxed{3}ask(a, b, ?.leave(x)) \\ \text{CONTEXT} & \begin{bmatrix} \text{SAL-UTT} & \begin{bmatrix} \text{PHON} & \langle \text{bo} \rangle \\ \text{CONTENT} & \boxed{1}\begin{bmatrix} x\!:\!name(x, Bo) \end{bmatrix} \end{bmatrix} \\ \text{MAX-QUD} & ?\boxed{1}.\boxed{3}ask(a, b, ?.leave(x)) \end{bmatrix} \end{bmatrix}$$

Constituent Readings   In the case where the hearer can find no value for a parameter in context, the question that arises is a *constituent* question, a *wh*-question about the intended content of the problematic utterance, corresponding to the example and paraphrase given here as example (28).

(28)
	A:	Did Bo leave?
	B:	**BO?**
	A:	Bo Smith.
	B:	Yes, half an hour ago.
	$\rightsquigarrow$	*"What is the intended content of your utterance 'Bo'?"*

For this reading, the coercion operation results in an updated context where the maximal QUD is precisely this question about the intended content of the sub-utterance, *"Who do you mean by 'Bo'?"*, or more specifically *"Which individual did you intend to be the content of*

---

[12]There are also syntactic parallelism constraints on SAL-UTT which are not shown here.

your utterance 'Bo'?", $?x.spkr\_meaning\_rel(a, \text{`Bo'}, x)$.

(29)

$$\begin{bmatrix} \text{C-PARAMS} & \left\{ \ldots, \boxed{1}, \ldots \right\} \\ \text{CONSTITS} & \left\{ \ldots, \boxed{2}\left[\text{CONTENT} \quad \boxed{1} \right], \ldots \right\} \end{bmatrix}$$
(original utterance)

$\Rightarrow$

$$\begin{bmatrix} \text{CONTEXT} & \begin{bmatrix} \text{SAL-UTT} & \boxed{2} \\ \text{MAX-QUD} & ?\boxed{1}.spkr\_meaning\_rel(a, \boxed{2}, \boxed{1}) \end{bmatrix} \end{bmatrix}$$
(partial reprise context description)

In this case the elliptical question *"Bo?"* must be assigned an *utterance-anaphoric* analysis by the grammar. They implement this via a specific phrase type, *utt(erance)-anaph(oric)-ph(rase)*, as shown in AVM (30), which enables reference to a previous (sub-)utterance. This phrase type takes a fragment as its only daughter (restricted, as is the case in all of G&C's analyses, to be a NP), but does not derive its semantic content from that daughter: instead, content is a parameter whose referent is a salient utterance in context, which has the same phonological form as the daughter. In this way, an utterance-anaphoric word *Bo* would refer to a salient contextual utterance of the word *Bo*.

(30)

$$\begin{bmatrix} \textit{utt-anaph-ph} \\ \text{CONT} & \boxed{1}:\boxed{1} = \boxed{2} \\ \text{HEAD-DTR} & \begin{bmatrix} \text{CAT} & \textit{NP} \\ \text{PHON} & \boxed{3} \\ \text{CTXT} \mid \text{SAL-UTT} & \boxed{2}\left[\text{PHON} \quad \boxed{3} \right] \end{bmatrix} \end{bmatrix}$$

The fragment *"Bo?"* (delivered with suitable intonation), is now analysed using a constituent-CR-specific phrase type which requires an utterance-anaphoric phrase as its daughter and assigns semantic content directly from the contextual MAX-QUD:

(31)

$$\begin{bmatrix} \textit{constit-clar-int-cl} \\ \text{CONT} & \boxed{1} \\ \text{HEAD-DTR} & \left[\textit{utt-anaph-ph} \right] \\ \text{CTXT} \mid \text{MAX-QUD} & \boxed{1} \end{bmatrix}$$

Combining these constraints with the new updated context forces this partial specification to be fully resolved as having the new MAX-QUD question as its content:

(32)

$$\begin{bmatrix} \text{CONTENT} & ?\boxed{1}.spkr\_meaning\_rel(a, \boxed{2}, \boxed{1}) \\ \text{CONTEXT} \mid \text{SAL-UTT} & \boxed{2}\begin{bmatrix} \textit{sign} \\ \text{PHON} & \left\langle \text{bo} \right\rangle \\ \text{CONTENT} & \boxed{1}\left[x\!:\!name(x, Bo) \right] \end{bmatrix} \end{bmatrix}$$

Other Possible Readings   A possible *lexical identification* reading is also discussed and taken to be consistent with the utterance-anaphoric approach, although no analysis is given:

this is a question concerning surface form (phonology or orthography) of the words used by the speaker (for example, in situations with high background noise levels), rather than semantic content. This corresponds to what (Iwata, 2003) calls the metalinguistic reading (metarepresenting form), as opposed to the metaconceptual (metarepresenting content), and takes to be required for questions about pronunciation and sub-lexical questions like example (5).

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.

2.3.6   Summary

G&C's analysis, then, provides many of the features desired here: a model potentially explaining how CRs arise in dialogue and what causes them; a well-defined grammatical method of building up syntactic and semantic analyses (of elliptical forms as well as full sentences) that does not rely on heavyweight inference; and the possibility of analysing several different syntactic forms and semantic readings. It also treats CRs as standard interrogatives, whose content is a standard question (which happens to be about a previous utterance), and this seems advantageous if the overall approach is to be integrated into a general theory of dialogue. Of course, we do not know what other readings and forms there may be, or how realistically implementable the grammar is. Also, and importantly, the approach is only defined for proper names and full sentences, and real clarificational dialogue is clearly not restricted to these.

G&C use HPSG for their analysis, and the extensions provided in the following chapters will follow this. Although they believe that the analysis is applicable to other frameworks, HPSG provides certain features that are advantageous when dealing with reprise questions: in particular, direct access to phonological, syntactic, semantic and contextual information and the availability of constraints between these levels; and the ability to treat utterances as objects within the grammar.[13]

2.4   Computational Approaches

Designers of computational dialogue systems have long been aware of the need for the system to be able to clarify user input in some way, to indicate to the user that the latest input was not properly perceived or understood. In most cases, this is as far as it goes: systems are capable of exactly this and this only, producing prompts such as *"I'm sorry, I do not understand.*

---

[13]For an alternative formulation of some of G&C's account within Martin-Löf Type Theory, see (Cooper and Ginzburg, 2002). (Poesio and Traum, 1997) also provide a DRT-based framework which includes utterance reference.

Please repeat." Some have gone further, however, showing that system-generated clarification can help provide useful capabilities.

In contrast, there is a marked absence of work enabling systems to deal with user-generated clarification. The general approach in system design has always been to make system prompts as clear and unambiguous as possible, thus ideally preventing users from needing to initiate clarificational dialogue at all. However, while this has proved a reasonable approach for systems which work with a limited domain, a limited population of users with limited expectations of the system's capabilities, or both, it may become less viable as systems become more human-like and have to deal with wider domains and user pools.

This section introduces the system that is used as the baseline for the implementation of chapter 6 and describes its approach to clarification, which is representative of the current state-of-the-art system. It then describes some other approaches to and uses of clarification that have been taken in the computational dialogue system field, and illustrates why an integrated treatment of clarification might provide real benefits.

2.4.1 GoDiS/IBiS

The GoDiS system (Larsson et al., 2000) and its successor IBiS (Larsson, 2002) give a good picture of a typical approach to clarification. The basic GoDiS system recognises the need for CRs to allow the system to indicate when interpretation has failed, but only two possible questions can be produced: *"I didn't understand what you said. Please rephrase."*, which is generated when the interpretation process fails entirely, producing no semantic representation for the most recent user turn, and *"What do you mean by that?"*, which is used when semantic interpretation succeeds but the move that appears to have been made is not relevant and cannot be successfully processed given the current context and knowledge of the system domain.

In IBiS, the system CR capability is more complex and specifically reflects (Allwood, 2000)'s levels of communication, with system CRs available to query each level: *"Pardon?"*, *"What did you say?"*, *"Sorry, I didn't hear what you said."* to query at the perception level; *"What do you mean?"*, *"Sorry, I don't understand."* to query at the level of semantic understanding; and *"I don't quite understand."* at the level of pragmatic understanding (relevance to the current context). At this level, a form of reprise question is also possible: user sentences can be echoed back to check whether the system's interpretation of them was right, e.g. *"You asked about price information, is that correct?"* – see listing 2.1.[14]

In both systems, user CRs are possible, but limited to *"What did you say?"*, *"What?"*, *"Pardon?"* and *"Sorry?"*, all of which are taken to be asking the system to repeat its entire last utterance (see listing 2.2)[15] – there is no way a user can ask about a particular word or

---

[14]From Larsson (2002)'s dialogues (3.4), p.114 and (3.5), p.120. The use of a different noun in the echo (*flight* vs. *plane*) is not a deliberate strategy – both are synonyms for the same concept, for which the system will always generate *flight*.

[15]Larsson (2002)'s dialogue (3.10), p.129

```
U> jfdbhajhdgarbledfsasd
S> I heard you say jfdbhajhdgarbledfsasd. Sorry, I don't understand.
...
U> paris
S> paris. I don't quite understand.
...
S> How do you want to travel?
U> by plane
S> by flight, is that correct?
U> no
S> not by flight. So, how do you want to travel?
```

Listing 2.1: Sample IBiS System CRs

part of an utterance, or ask any questions about semantic content or reference.

```
S> Okay. You asked about price. I need some information. How do you want
   to travel?
U> pardon
S> Okay. You asked about price. I need some information. How do you want
   to travel?
```

Listing 2.2: Sample IBiS User CR

The framework of IBiS would allow a greater range of clarificational dialogue to be handled in principle: like Traum (1994), its main area of research has been in modelling the mechanism of grounding and acceptance (positive or negative) and how the dialogue context evolves, rather than the clarificational dialogue which might ensue. Its information-state approach allows the kind of access to context which will be required for a treatment of clarification. However, as will become clear, some fundamental changes will be required to the way information is represented.

CRs are seen as moves of a distinct type from standard questions (in Larsson's terminology, `icm` or *interactive communication management* moves rather than `ask` moves). This has the advantage of allowing them to have different effects (e.g. following (Clark, 1996), they are not taken to require explicit grounding themselves), but has the disadvantage that they do not necessarily follow the standard rules of dialogue, e.g. raising QUDs and thereby licensing elliptical answers.

2.4.2 Recognition Problems

There are several ways in which a more advanced approach to clarification might prove useful. IBiS uses confidence scores produced by its speech recognition module to decide whether an utterance has been acceptably perceived or not: below a certain score, the entire utterance is echoed back and the system asks *"Is that correct?"*. A more sophisticated approach to could

allow CRs to ask about individual problematic parts rather than the whole utterance, making them clearer and more helpful to the user. Gabsdil (2003) and Gabsdil and Bos (2003) propose using speech recognition confidence scores to pinpoint the source of the problem more accurately, identifying the word(s) or semantic sub-formulae with the lowest confidence scores and thereby generating a CR which explicitly queries these parts. As they acknowledge, however, such CRs can take many forms, and generating them is far from trivial, especially in a system which uses a complex semantic representation (rather than simple slots and fillers). As will become apparent later, the ability to generate these various CR forms may be important as different forms may be associated with different types of source problem and with different answer expectations.

2.4.3 Unknown Words

Robust keyword-spotting approaches and current speech recognisers deal with out-of-vocabulary words by ignoring them or recognising them as known in-vocabulary words. This can cause problems when user utterances are rejected (or have the wrong effect) without giving the user any indication of which word has caused the problem and why. The RIALIST (Hockey et al., 2002) and On/Off House (Gorrell et al., 2002) systems take a step towards dealing with this by using two speech recognisers: a main grammar-based recogniser which has high accuracy but a limited grammar and vocabulary, and a second statistical recogniser which is less accurate but has a much wider vocabulary. When the main recogniser gives a very low confidence score, and the backup recogniser instead gives a high confidence score in a string which contains a word not in the main lexicon, the system can produce a prompt which tells the user which this word is, and gives an example of what an acceptable utterance might be in the current context (see example (33) below).

(33)
User:	Go to unit one.
System:	OK. Now at unit one.
User:	Measure the humidity.
System:	I'm sorry, I don't understand the word 'humidity'. Please rephrase - for example, you could say "measure the carbon dioxide level".

Compared to rejecting the utterance outright (as most systems would do) this does seem a significant improvement as it helps the user reformulate their utterance in a successful way. However, the user is forced to go back and reformulate their utterance completely – there is no way that the meaning of the new word can be discussed and determined. In fact, as this prompt is produced by a separate dedicated module (outside the normal dialogue process), it would be very difficult for this to be achieved. Also, there is no way for the system to learn the word, and it will cause exactly the same problem if it appears again later in the dialogue.

So an approach which can process and ask about unknown words seems very useful, but in an ideal world would be incorporated within the standard dialogue process, and would

enable words to be discussed and learnt.

2.4.4   Lexicon and Grammar Acquisition

Learning new words within a dialogue system is a difficult problem. Unknown words must be processed on their first appearance, with the only contextual information available being that given by the surrounding sentence. This rules out the experience-based approaches common in automatic acquisition (e.g. Pedersen, 1995; Barg and Walther, 1998), which need large amounts of data to acquire any semantic information beyond broad category and argument selection information. This local context may be sufficient to allow syntactic information to be inferred (at least sufficient to parse the sentence), but nothing beyond this.

One possible approach is knowledge-based: the FOUL-UP system (Granger, 1977) used scripted information about a known situation to infer semantic information about words. This approach can gain detailed information about words which play major roles in sentences, although very little progress can be made with modifiers e.g. adjectives. More importantly, the domain is limited, and the approach cannot be applied to a system intended for open- or wide-domain use.

The second approach is of course to depend on user interaction. This has been used in standard text-processing systems such as VEX in the Core Language Engine (Carter, 1992) which asks the user to select from a list of possible usages. Within a dialogue system, however, clarificational dialogue must be used. There are precedents: the RINA system (Zernik, 1987) used questions about the meaning of phrases in a system simulating a second language learner, Knight (1996) proposes the use of questions about word meaning in a machine translation system, and more recently Dusan and Flanagan (2001, 2002) use a version of the dual-recogniser approach described above together with clarification questions within a multimodal dialogue system. However, these systems understandably treat the clarification exchange as self-contained and governed by its own rules; but within a standard dialogue system it cannot necessarily be distinguished from the wider dialogue. Ideally, such clarificational dialogue would be governed by standard rules and thus subject to the standard processes of ellipsis and anaphora resolution, answer recognition, and indeed further clarification.

Rosé (1997) goes further, extending this kind of approach from individual words to whole phrases and sentences, in the robust interpretation system ROSE. Faced with extragrammatical input, it can make hypotheses about possible sentence interpretations, and then rely on user interaction to discover which hypothesis (if any) is correct. This interaction has two possible styles, either asking explicit questions about sentence meaning with a set form such as *"Was something like X part of what you meant?"* where X is a sentence or phrase, or more natural but task-dependent questions about intended actions, such as *"Are you suggesting that X is a good time to meet?"* for a scheduling task. Again, clarification is thus restricted to particular phenomena and forms, and is a self-contained process rather than having to be

recognised and integrated as part of a dialogue.

2.4.5 Reference Resolution

A further motivation for system-initiated clarification is the possibility of ambiguity, and in particular the common ambiguity of reference of referring expressions such as definite descriptions and pronouns. Heeman and Hirst (1995) describe an intentional plan-based method of discussing the referents of such expressions when they cannot be found. More recently, Traum (2003) describes a multi-agent system in which agents are capable of producing CRs intended to clarify this reference and then using the response to fully specify the previous utterance. In example (34), the user is a human trainee (a military officer) and the system plays the part of two agents, a sergeant and a medic, in a scenario intended to rehearse peacekeeping missions in Bosnia. The clarification request is shown in **bold**, with the problematic source and the answer both shown <u>underlined</u> (these conventions will be used throughout):

(34)

	User:	What happened here?
	Sergeant:	There was an accident sir
	User:	Who's hurt?
	Sergeant:	The boy and one of our drivers
	User:	How bad is <u>he</u> hurt?
	Sergeant:	**The driver or the boy?**
	User:	<u>The boy</u>
	Sergeant:	Tucci?
	Medic:	The boy has critical injuries
	Sergeant:	Understood

This is much closer to the kind of human-like behaviour that we are working towards: the original utterance does not have to be repeated, the clarification question is natural (and elliptical), and the answer can be processed and the dialogue continues as normal. The process of initiating clarification also fits within the general rules of dialogue, based in the grounding approach of (Traum, 1994; Poesio and Traum, 1998), and as a result this appears to be one of the few systems that can accept user-initiated CRs. However, there are limitations: firstly, the system is currently limited to this type of clarification of a nominal referring expression, and a general request for repair of an incomprehensible utterance *"Say again?"* (although the approach could certainly be extended to other phenomena in theory); secondly, as with GoDiS and IBiS, the CR is taken to be a specific dialogue act `request-repair` which has rather different effects from asking a normal question.[16] For example, CRs do not introduce a question to QUD as other questions do (this being the standard mechanism for answer recognition and fragment resolution), or introduce content to the context in the same way (thus complicating any account of how CRs might become the subject of clarification themselves,

---

[16]Similar observations apply to the model defined in (Heeman and Hirst, 1995), which is apparently limited to suggesting expanded descriptions of nominal referents, realised as a particular speech act.

which as we will see does appear to happen). So again, examining a more general approach seems beneficial, as does studying the possible different types of clarification.

2.4.6 Summary

While the requirement for systems to produce CRs is undisputed, capabilities are usually restricted to general high-level indications of lack of comprehension. The motivation for allowing systems to produce more targeted, detailed and useful CRs is clear: indeed, there are as many reasons why they are required or advantageous as there are types of information that might be clarified – spoken word identification, word meaning identification and acquisition, and reference resolution being amongst them. These different problems have been treated by various people in various ways, most of which are idiosyncratic and rely on specially defined routines and/or dialogue acts. What appears to be needed is an integrated approach which can take in the different types of CRs and can fit into general dialogue processing.

Furthermore, user-initiated clarification has received very little attention in the field (Traum (2003) being an honourable exception), and studying clarification and its various forms and readings seems an important task, with a view to building systems that can correctly interpret and participate in such dialogue.

2.5 GoDiS, the TrindiKit and SHARDS

This section gives a brief overview of the GoDiS system (Larsson et al., 2000) and the TrindiKit framework and toolbox (Larsson et al., 1999), as well as the SHARDS ellipsis reconstruction system (Ginzburg et al., 2001a; Fernández, 2002; Fernández et al., 2004a), which are used as starting points for CLARIE, the dialogue system prototype which will be described later in chapter 6.

2.5.1 The GoDiS System

System Overview

GoDiS, introduced in the previous section, is a dialogue system implemented in Prolog using the TrindiKit framework, and which is based on KOS and the QUD approach to dialogue modelling (Ginzburg, 1996, forthcoming) – as such it is ideal as the starting point for an implementation based on G&C's QUD-based analysis. This section gives some background on the basic principles of the system and the framework used (in particular the information state model and how it is updated) – which will be used as the basis of the system in chapter 6. Those familiar with the TrindiKit may want to skip this section.

The system centres around an *information state* (IS – see e.g. Cooper et al., 1999), a structured representation of the context and the state of the dialogue and the system at any

point. The TrindiKit allows a modular structure: various modules can feed information into the IS (such as the latest utterance from the user), or can read information from it (such as the next utterance that the system should produce), and a central *dialogue move engine* (DME) updates the IS after each utterance using a set of update rules. This structure is shown in figure 2.2 (taken from Larsson, 2000).
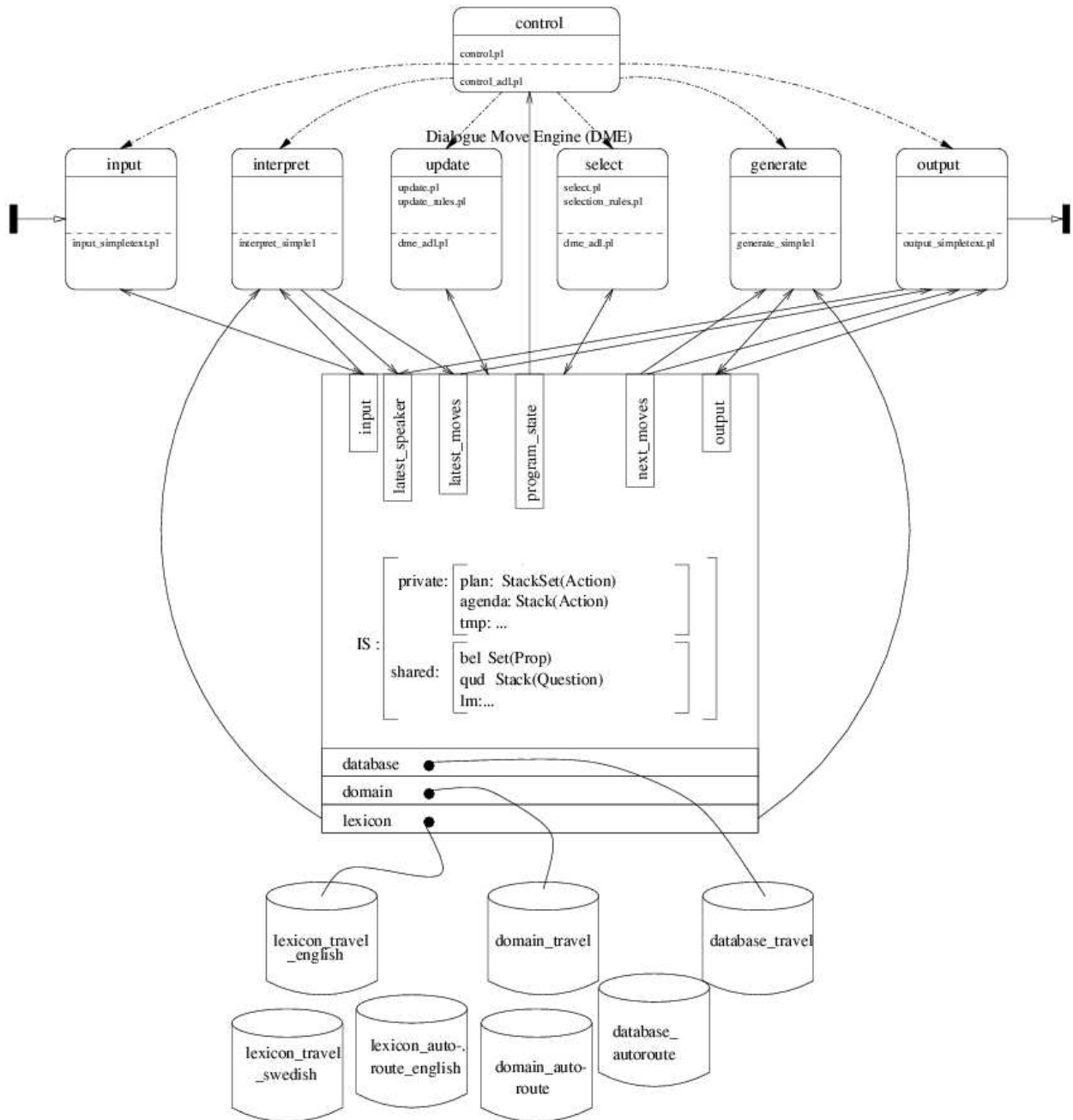
Figure 2.2: GoDiS System Structure

The modules that make up GoDiS are as follows:

- `input` – this module takes the user input (a text string or speech signal) and provides a corresponding list of words;

- `interpret` – this module converts the list of words into meaning, a set of *dialogue moves* (the actions intended to be performed by the input, e.g. asking a particular question, answering a previous question);

- `update` – this forms the first part of the DME, and defines how the IS is updated given the dialogue moves made, including updating the system's immediate intentions;

- `select` – the second part of the DME, this defines what dialogue moves should next be made by the system, given the IS update;

- `generate` – this module converts the set of system dialogue moves into a corresponding text string, the inverse of the `interpret` module;

- `output` – and the string is converted into output text or speech;

- `control` – finally, the control module maintains the overall process, calling each other module in turn.

The TrindiKit also allows *resources* to be defined, modules which do not play a particular part in the overall control algorithm but which provide particular information or capabilities which can be called upon by other modules. In GoDiS, three are defined:

- `domain` – containing information and knowledge related to a specific domain, e.g. departure points, countries and destinations for its travel agent implementation;

- `lexicon` – containing information relating words to (domain-specific) meanings;

- `database` – performing calculations of ticket prices, allowing certain questions to be answered.

This modular nature allows domain and language to be changed easily, and also allows individual modules to be replaced with new versions which use different methods or strategies; in CLARIE's case, both interpretation and generation modules will be replaced, and the DME rules will be adapted.

Information State

The GoDiS information state (IS) is shown in AVM (35). It consists of two parts, following (Ginzburg, 1996), a PRIVATE part associated with information available to the system

but as yet unpublicised in the dialogue, and a SHARED part for information which has been publicised.

$$
(35)\quad
\begin{bmatrix}
\text{PRIVATE} &
\begin{bmatrix}
\text{AGENDA} & \begin{bmatrix} stack(action) \end{bmatrix} \\
\text{PLAN} & \begin{bmatrix} stackset(action) \end{bmatrix} \\
\text{BEL} & \begin{bmatrix} set(proposition) \end{bmatrix} \\
\text{TMP} & \begin{bmatrix} shared \end{bmatrix}
\end{bmatrix} \\[4em]
\text{SHARED} &
\begin{bmatrix}
\text{COM} & \begin{bmatrix} set(proposition) \end{bmatrix} \\
\text{QUD} & \begin{bmatrix} stack(question) \end{bmatrix} \\
\text{LU} & \begin{bmatrix}
\text{SPEAKER} & \begin{bmatrix} speaker \end{bmatrix} \\
\text{MOVES} & \begin{bmatrix} assocset(move,bool) \end{bmatrix}
\end{bmatrix} \\
\text{NIM} & \begin{bmatrix} stackset(move) \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

The private part of the IS is used by GoDiS mainly for plan storage and management: the PLAN record holds the current overall plan (questions that must be raised and answered, information that must be given to the user etc.), while the AGENDA record holds the action immediately under consideration (e.g. the question that has just been raised or will be raised as soon as possible). There is also a BEL record for privately held beliefs (with the travel agent plan, this is used to store information like ticket price that has been determined by the system but not yet given to the user). The final private record is called TMP, and is used as a temporary storage mechanism to allow backtracking: a copy of the SHARED slate is kept at each turn so that if the next turn reveals that the user has not understood a system utterance, the system can revert to the old state. This is required in GoDiS due to the combination of the optimistic grounding strategy that is taken and the lack of an utterance record beyond the current turn: system utterances are assumed to be understood by the user and are added to the common ground as they are output. This therefore necessitates backtracking if the next user move is to request a repeat (*"Pardon?/What did you say?"* is the only user CR that GoDiS allows).

The shared part of the IS concerns the information that results from the actual dialogue, as it has happened up until any given point. Again following (Ginzburg, 1996), the COM record holds shared commitments, a set of propositions that have been established in the common ground, and the QUD record is a stack[17] of questions under discussion. Information about the latest utterance is stored in the LU record, which contains two sub-records: SPEAKER to identify the latest speaker (system or user); and MOVES, containing the dialogue moves made

---

[17]The TrindiKit provides certain common data types which are used in both GoDiS and CLARIE: a *stack* is an ordered array of which only the top element is accessible; a *set* is an unordered array of which any element is accessible; a *stackset* has the ordering of a stack with the accessibility of a set; an *assocset* is a set paired with a second set of flags, with each element of one set associated with a distinct element of the second.

by the utterance, each associated with a boolean flag that records whether it has been grounded and integrated into the IS. The final record, called NIM (for Non-Incorporated Moves), is used for part of the grounding process to deal with moves whose pragmatic relevance to the current plan cannot be established (although the associated utterance has been understood semantically as making the moves).

Interpretation and Generation

The interpretation module works by spotting keywords and phrases and interpreting them in a domain-specific manner, with the lexicon and domain resources together specifying the possible keywords and their interpretations, as shown in listing 2.3:

```
input_form( [to|S], answer(to(C)) ) :- lexsem(S,C), location(C).
input_form( [from|S], answer(from(C)) ) :- lexsem(S,C), location(C).
input_form( [by|S], answer(how(C)) ) :- lexsem(S,C), means_of_transport(C).
input_form( [price], answer(task(price_info))).
input_form( [reservation], answer(task(order_trip))).

...
location( paris ).
...
month( march ).
...
```

Listing 2.3: Sample GoDiS input templates

The resulting interpretation is a set of conversational moves. These can be answers to particular questions as shown in listing 2.4 below, can be asking questions or asserting propositional information, can be greeting or closing moves, or can be requests for repetition (the clarificational move type as described above).

```
User> i want to go to paris in march please

:   latest_speaker = usr
:   latest_moves = { answer(to(paris)), answer(month(march)) }
```

Listing 2.4: Sample GoDiS interpretation

This method has the advantage of being extremely robust (no parsing problems will be caused by noisy or ungrammatical input), but is of course highly domain-specific and restricted.

Similarly, the generation module produces canned text output, predefined in the lexicon as corresponding to each of the dialogue moves that the system can make:

```
output_form( greet, "Welcome to the travel agency!" ).
output_form( quit, "Thank you for your visit!" ).
output_form( ask(return), "Do you want a return ticket?" ).
output_form( ask(X^(from(X))), "What city do you want to go from?" ).
output_form( ask(X^(to(X))), "What city do you want to go to?" ).
```

Listing 2.5: Sample GoDiS output templates

Dialogue Management

The strength of GoDiS lies in its IS representation and ability to use all the information this contains in interpreting moves and deciding on suitable responses. This process is performed by the first part of the DME (the `update` module), which is characterised by an overall update algorithm together with a set of individual rules which can be applied if particular contextual conditions hold. The update algorithm (or a simplified version thereof) is shown in listing 2.6:

```
if ( $latest_moves == failed )
then ([ repeat refill_agenda ])
else ([ if ( $latest_speaker == sys )
        then ([ try integrate,
                try database,
                repeat downdate_agenda,
                store ])
        else ([ repeat ( integrate or
                         accommodate or
                         find_plan ),
                repeat downdate_agenda,
                repeat manage_plan,
                repeat refill_agenda,
                repeat store_nim,
                try downdate_qud ]) ])
```

Listing 2.6: GoDiS update algorithm

Without delving too far into the details, the basic process goes as follows. If the `latest_moves` variable (containing the set of moves assigned to the latest utterance by the interpretation module) shows that the last utterance could not be interpreted at all, the system ignores it, goes ahead with its plan and puts the next item on the agenda. Otherwise it processes the latest move normally: if it was produced by the system, it will integrate the move into the IS using the `integrate` rules, then perform a number of tidying-up jobs (seeing if it has enough information to look up a ticket price in its database, removing any agenda items that have been resolved, and storing the current IS in the TMP field for possible later backtracking). The integration rules define all the immediate effects that particular moves have on the IS. If on the other hand the utterance was produced by the user, it will go through an iterative

process of trying to integrate the moves into the IS as they stand; then if unsuccessful, trying to *accommodate* a suitable question from the current plan such that integration is possible, or even trying to find another plan such that accommodation or integration can succeed. Then a similar tidying-up process begins.

It is therefore the integration rules that define the behaviour of the system given a particular move: how the IS is updated governs how it responds to questions and answers, how it builds up beliefs and commitments, and how it carries out its overall plan. The accommodation rules allow more adaptive and robust behaviour by allowing moves to be interpreted and integrated as if planned (but not actually asked) questions had been in context. It is these two sets of rules that make up the body of the update process.

DME Update Rules

All DME rules are specified as TrindiKit update rules, which have a specific syntax. An update rule consists of three parts: the name of the rule, a list of conditions that must be met for the rule to apply, and a list of effects that will be carried out when it is applied. Details of the syntax are available in (Larsson et al., 1999), but a summary is as follows: IS fields are specified using the Unix-like / operator; the $ operator addresses *values* of fields rather than the fields themselves; modules and resources are addressed using a $module:: prefix; and each datatype has a defined set of possible operators (including the familiar push and pop for stacks, in for set membership, and so on).

```
rule( integrateSysAsk,
      [ $/shared/lu/speaker == sys,
        assoc( $/shared/lu/moves, ask(Q), false ),
        fst( $/private/agenda, raise(Q) ) ],
      [ push( /shared/qud, Q ),
        pop( /private/agenda ),
        set_assoc( /shared/lu/moves, ask(Q), true ) ] ).
```

Listing 2.7: GoDiS integration rule (system ask)

A typical rule for a system utterance is shown in listing 2.7 above: a rule for integrating a system ask move into the IS. In this case, the conditions check that the system was the speaker of the latest move, and that this move asked a question Q which was on the agenda to be raised. The effects push Q onto the stack of QUDs, remove the now-fulfilled agenda action, and set a flag showing that the move has now been integrated.

Integration Rules   The rules for integrating user moves are similar but also include the required response of the system. A rule for integrating a user ask move is shown in listing 2.8 – the asked question Q is pushed onto the QUD stack, and an action to respond to Q is pushed

onto the agenda (now becoming the top action, which will therefore be the first to be carried out):

```
rule( integrateUsrAsk,
      [ $/shared/lu/speaker == usr,
        assoc( $/shared/lu/moves,  ask(Q), false ) ],
      [ set_assoc( /shared/lu/moves, ask(Q), true ),
        push( /shared/qud, Q ),
        push( /private/agenda, respond(Q) ) ] ).
```

Listing 2.8: GoDiS integration rule (user answer)

A similar rule for integrating an answer move is shown in listing 2.9 below: in this case, it checks that it is an answer R to the question Q that is currently first on the QUD list, and which has not already been answered (i.e. there is no proposition $P1$ in the shared commitments that answers Q). It also checks that the question Q and answer R produce a full proposition P by a process of beta-reduction (defined as the reduce/3 condition). The effects are to flag the move as integrated, then to remove the answered question from QUD and to add the new proposition P to the shared commitments.

```
rule( integrateUsrAnswer,
      [ $/shared/lu/speaker == usr,
        assoc( $/shared/lu/moves, answer(R), false ),
        fst( $/shared/qud, Q ),
        $domain :: relevant_answer(Q,R),
        not( in( $/shared/com, P1 ) and $domain :: relevant_answer(Q,P1) ),
        $domain :: reduce(Q,R,P) ],
      [ set_assoc( /shared/lu/moves, answer(R), true ),
        pop( /shared/qud ),
        add( /shared/com, P ) ] ).
```

Listing 2.9: GoDiS integration rule (user answer)

This rule also has a side-effect of performing simple ellipsis resolution: an answer of just paris, rather than to(paris) or from(paris), can be resolved by the relevant_answer predicate as meaning one or the other depending on whether the current QUD Q is the question $?x.to(x)$ (*"Where do you want to go to?"*) or $?x.from(x)$ (*"Where do you want to go from?"*).

Accommodation Rules   The accommodation rules allow a question from the agenda or plan to be used to resolve answers as above, even when they have not been explicitly asked. This allows a user to give more information than is requested (e.g. when asked *"Where do you want to go?"*, to reply *"To Paris, from London in March"*). The extra information can then be seen as the answer to a planned question; this question is then removed from the plan and

instead added to QUD, and now the standard integration rule can apply:

```
rule( accommodateQuestion,
      [ $/shared/lu/speaker == usr,
        in( $/shared/lu/moves, answer(A) ),
        not( $lexicon :: yn_answer(A) ),
        assoc( $/shared/lu/moves, answer(A), false ),
        in( $/private/plan, findout(Q) ),
        $domain :: relevant_answer(Q,A) ],
      [ del( /private/plan, findout(Q) ),
        push( /shared/qud, Q ) ] ).
```

Listing 2.10: GoDiS accommodation rule

Selection Rules

The second part of the DME is the selection process, which control the output process just as
the update rules control the interpretation process. The TrindiKit syntax is the same, and the
rules themselves are generally much simpler, translating a planned action into a corresponding
move (which subsequently forms the input to the generation module, which turns it into a
sentence). Selection rules all take the form of the example in listing 2.11 below, which governs
the selection of answer moves. If the action is to respond to a question Q, and if there is a
belief R in the IS which answers Q, then the next move will be to answer with R:

```
rule( selectAnswer,
      [ fst( $/private/agenda, respond(Q) ),
        in( $/private/bel, R ),
        $domain :: relevant_answer(Q,R) ],
      [ set( next_moves, set([ answer(Q,R) ]) ) ] ).
```

Listing 2.11: GoDiS selection rule

Similar rules apply to relate each of the possible system actions (respond, findout,
quit etc.) to corresponding dialogue moves (answer, ask, quit).

The basic GoDiS system therefore provides a basic framework which can be easily ex-
tended by altering the various modules as required, together with a flexible approach to dia-
logue management which allows integration and selection rules to interact with IS features as
required. Chapter 6 takes up the task of extending the capabilities to clarificational dialogue,
which will require new approaches to interpretation, generation and dialogue management.

2.5.2 The SHARDS System

The approach to ellipsis resolution assumed (and extended) by G&C has also been implemented in Prolog[18] as the SHARDS system. While not a dialogue system as such, SHARDS incorporates a simple model of dialogue context and uses this along with a HPSG grammar to resolve certain elliptical forms. This grammar and general approach to ellipsis is used as the basis for the grammar of chapter 5, which is then used in the interpretation and generation modules of the system of chapter 6, so this section gives a description; readers familiar with SHARDS or QUD-based approaches to ellipsis may want to skip this section.

The SHARDS grammar produces a sign representation which, in the case of elliptical fragments, has an underspecified semantic content which is dependent on certain contextual features. This underspecified sign is passed to an ellipsis reconstruction module, which uses contextual information (a set of possible questions under discussion (QUDs) and a set of possible salient utterances (SAL-UTTs)) to instantiate these features and fully specify the sign and its content.

The baseline SHARDS system is capable of processing short answers (example (36)), polar answers (example (37)) and direct sluices (example (38)).

(36)
> A: Who left?
> B: **John.**
>
> $\leadsto$ *"John left."*

(37)
> A: Did Mary leave?
> B: **Yes/Probably.**
>
> $\leadsto$ *"It is (probably) the case that Mary left."*

(38)
> A: A girl left.
> B: **Who?**
>
> $\leadsto$ *"Which girl left?"*

Note that direct sluices as in example (38) are not quite the same as the reprise sluices described in section 2.3.5 above. Reprise sluices are CRs – they ask about the intended content of the previous utterance, which has not been successfully grounded or fully understood. Direct sluices are not – they ask for further information than was actually provided by the previous utterance, which may have been understood perfectly.[19]

---

[18]SHARDS uses ProFIT (Erbach, 1995) to represent HPSG feature structures in Prolog.

[19]Recent results (Fernández et al., 2004b) show that these two types of sluice can be distinguished by human judges with good cross-annotator agreement, and by machine learning techniques.

Resolution (Short Answers)

The grammar assigns the underspecified, context-dependent content as follows. For short answers, content is a proposition determined by the current maximal QUD, which must be the value of the contextual feature MAX-QUD (as shown in AVM (39) for example (36)). A second contextual feature SAL-UTT also links the index of the fragment (its individual referent) with that of a salient utterance in context, while expressing certain syntactic constraints which are left out here for simplicity:

$$(39) \quad \begin{bmatrix} \textit{decl-frag-cl} \\ \text{PHON} & \left\langle \text{john} \right\rangle \\ \text{CONT} & \boxed{1}\begin{bmatrix} \textit{proposition} \end{bmatrix} \\ \text{HEAD-DTR} \mid \text{CONT} & \begin{bmatrix} \boxed{2} : name(\boxed{2}, john) \end{bmatrix} \\ \text{CTXT} & \begin{bmatrix} \text{MAX-QUD} \mid \text{PROP} & \boxed{1} \\ \text{SAL-UTT} \mid \text{CONT} \mid \text{INDEX} & \boxed{2} \end{bmatrix} \end{bmatrix}$$

The ellipsis reconstruction module then uses the current possible values of MAX-QUD and SAL-UTT, which are calculated from the dialogue context, to fully specify the sign. The possible values are calculated using simple dialogue processing rules as follows:

1. Asking any question q raises q as MAX-QUD.

   – If q is a *wh*-question $?x.p$, then the constituent associated with x is made SAL-UTT.

2. Asserting any proposition p raises the question $?.p$ as MAX-QUD.

   – If p is an existentially quantified proposition $\exists x.p'$, then the constituent associated with x is made SAL-UTT.

In example (36) then, the first of these rules applies, and the value of MAX-QUD will be the question $?x.leave(x)$ (*"Who left?"*). The content of the elliptical fragment will therefore become the proposition $leave(x)$, as shown in AVM (40). The SAL-UTT feature specifies the value of this x: the salient utterance (in the original question) will be the word *who*; its index value is x; unifying this x with $john$ therefore ensures that the complete content is the

proposition $leave(john)$ (*"John left"*) as desired.

(40)
$$
\begin{bmatrix}
\textit{decl-frag-cl} \\
\text{PHON} \quad \left\langle \text{john} \right\rangle \\
\text{CONT} \quad \boxed{1} \\
\text{HEAD-DTR} \,|\, \text{CONT} \quad \left[\boxed{2}:name(\boxed{2}, john)\right] \\
\text{CTXT} \begin{bmatrix}
\text{MAX-QUD} \begin{bmatrix}
\textit{question} \\
\text{PARAMS} \quad \boxed{2} \\
\text{PROP} \quad \boxed{1}\left[leave(\boxed{2})\right]
\end{bmatrix} \\
\text{SAL-UTT} \begin{bmatrix}
\text{PHON} \quad \left\langle \text{who} \right\rangle \\
\text{CONT} \,|\, \text{INDEX} \quad \boxed{2}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

Other Fragment Types

Polar answers and sluices are assigned their underspecified grammatical content in similar ways: for polar answers, the content is also a proposition formed from the maximal QUD but modified by the adverbial relation given in the fragment itself, and SAL-UTT plays no role; for sluices, the content is a question again formed from the maximal QUD but with a new *wh*-parameter contributed by the sluice itself:

(41)
$$
\begin{bmatrix}
\textit{pol-frag-cl} \\
\text{PHON} \quad \left\langle \text{probably} \right\rangle \\
\text{CONT} \begin{bmatrix}
\textit{probable-rel} \\
\text{PROP} \quad \boxed{1}\left[\textit{proposition}\right]
\end{bmatrix} \\
\text{CTXT} \quad \left[\text{MAX-QUD} \,|\, \text{PROP} \quad \boxed{1}\right]
\end{bmatrix}
\qquad
\begin{bmatrix}
\textit{slu-int-cl} \\
\text{PHON} \quad \left\langle \text{who} \right\rangle \\
\text{CONT} \quad ?\boxed{2}.\boxed{1} \\
\text{HEAD-DTR} \,|\, \text{CONT} \,|\, \text{INDEX} \quad \boxed{2} \\
\text{CTXT} \begin{bmatrix}
\text{MAX-QUD} \,|\, \text{PROP} \quad \boxed{1} \\
\text{SAL-UTT} \,|\, \text{CONT} \,|\, \text{INDEX} \quad \boxed{2}
\end{bmatrix}
\end{bmatrix}
$$

The dialogue processing rules already given now provide all that is needed for these types. In example (37), the question *"Did Mary leave?"* is asked: the first rule therefore causes the corresponding question $?.leave(mary)$ to be raised as the maximal QUD. As shown in AVM (41) above, a polar fragment *"Probably"* will take its propositional content from this maximal QUD, so the final resolved content will therefore be the proposition $probable(leave(mary))$ (*"It is probable that Mary left"*).

In example (38), the assertion *"A girl left"* causes the MAX-QUD raised by the second rule to be a question paraphrasable *"Did a girl leave?"*, represented as something like $?.\exists x.girl(x) \land leave(x)$ (details of quantification aside). As the asserted proposition is existentially quantified, the sub-utterance *a girl* (with its associated INDEX value x) is made SAL-UTT. The sluice *"Who?"* can then be resolved as having the content $?x.girl(x) \land leave(x)$ as desired.

The SHARDS system therefore provides a framework for implementing the theoretical

approach of G&S: a basic HPSG grammar and processes for resolution of ellipsis given a model of context. The framework is extensible to other phenomena (see Fernández et al., 2004a, for a discussion of adjunct sluices), and in chapter 5 it is extended to include a grammar which covers clarification, which in chapter 6 is integrated into a GoDiS-like dialogue system and information-state-based contextual model.

Chapter 3

Empirical Observations

3.1 Introduction

This chapter describes the empirical evidence provided by corpus investigation and a series
of experiments into the nature of CRs as actually used in dialogue. In particular, it attempts
to address the following questions:

- What forms do CRs take, and what readings can these forms have?

- How common are CRs (and the various forms they can take)?

- When do CRs occur (what types of words and phrases cause them, and how long after-
  wards can a CR appear)?

- How do CR form and reading depend on the type of phrase being clarified?

- How and when are CRs answered?

3.1.1 Overview

Firstly, section 3.2 describes an attempt using a corpus of dialogue to classify the various
forms and readings that CRs can take, resulting in an ontology of CRs together with some
correlations between forms and readings. It also provides data on the observed distance be-
tween CRs and the utterance being clarified (and thus the required length of utterance memory
in a dialogue system).

Section 3.3 then adds further corpus data concerning the *sources* of CRs: which word
and phrase types are likely or unlikely to be the subject of clarification. It also investigates
correlations between CR form and reading and features of the word or phrase being clarified.

Section 3.4 presents corresponding corpus data on *responses* to CRs: the distance between
CRs and their answers, and relations between CR form and reading and likely response types.

In section 3.5, a new experimental technique is described together with the results of experiments which give further detailed data on one particular CR form, including information about the effect of the source word or phrase type on reading and answer type.

3.2   Corpus Investigation 1 – Ontology

This section describes an attempt to exhaustively categorise CR forms and readings based on corpus work, and discusses the implications of the results for possible grammatical analyses and for use in a practical system.[1] Taking G&C's work as a basic starting point, it is clear that there are several possible forms (at least reprise sentences, reprise fragments and reprise sluices) and more than one possible reading (they give semantic analyses for clausal and constituent questions, and mention the possibility of lexical form questions), but it is not clear whether all of the readings exist, whether all of the forms can take all of the readings, or what other forms and readings might exist.

The investigation also had a secondary aim. As typified by G&C's analysis, grammatical interpretation of CRs must 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). Other lower-level information such as propositional content or rhetorical structure may be kept longer, of course, and this may be required in dialogue systems (Moore (1993) points out that reference to previous discourse is common in tutorial dialogues, e.g. basing one explanation on another given previously), but keeping all levels indefinitely seems both costly and unrealistic. However, as shown in example (42), not all CRs (shown **bold**) come immediately after the utterance being clarified (the *source* utterance, shown <u>underlined</u>). This corpus work therefore had the additional aim of identification of the maximum *CR-source separation*

---

[1]Much of the work in this section has been published as (Purver et al., 2001, 2003a).

(CSS) distance between a CR and the source utterance.

$(42)^2$

Richard:	So er what do you think I should call my business?
Unknown:	`<unclear>`
Anon 5:	Richard
Unknown:	`<unclear>`
Anon 2:	I dunno, she's trying to
Unknown:	`<unclear>`
Anon 5:	`<laugh>`
Richard:	But erm
Anon 2:	**What you should call your business?**
Anon 6:	How do I know that that bath towel needs washing? `<pause>`
Anon 2:	**What shall you call your business?**
	Erm
Anon 6:	Ha Static Aquatic
Anon 5:	`<laugh>`

The next section 3.2.1 describes the corpus and methods used. In sections 3.2.2 and 3.2.3, the resulting CR forms and readings that were identified from corpus analysis are then listed. Section 3.2.4 gives detailed results, including a discussion of apparent correlations between certain forms and readings and of maximum observed CSS distance. Section 3.2.6 then discusses the implications of these findings for the intended dialogue system implementation.

3.2.1   Aims and Procedure

The intention was to investigate the forms, readings and CSS distances for CRs that are present in a corpus of dialogue. For this purpose we used the British National Corpus (BNC) (see Burnard, 2000), 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.

A total of 418 CRs within this sub-corpus were identified and tagged, using the markup scheme and decision process described below. The results given here are those produced by the first attempt, although the process has been repeated by a naive user to check its reliability, and 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 easy-to-read manner). However, in order to ensure that all clarificational phenomena were captured, the final search and markup were performed manually.

Corpus   The BNC includes many different types of dialogue from various domains, with transcripts being identified either as belonging to a particular context-governed domain (in-

---

[2]BNC file KSV, sentences 378–386

cluding business (meetings, training sessions), educational (school classes, lectures), and radio interviews – see (Burnard, 2000) for a full list), or as being *demographic* (general non-context-governed dialogue recorded by subjects during their daily lives. The majority of dialogues were taken from the demographic portion (90%), with the remainder from various domains. To maintain a spread across region, speaker age etc., the sub-corpus was created by taking a 200-speaker-turn section from 59 transcripts. Although the dialogue is recorded from natural speech, the BNC transcription itself does not include intonational markup; neither does it include any indication of facial expression, body language, gestures etc. which might be assumed to be common given that the majority of the dialogue is face-to-face. While most of the transcripts are (or appear to be) two-party dialogues, many involve more than two participants, and others (due to the demographic nature) may well involve other non-verbal participants or bystanders – this is discussed further in section 3.2.4.

This approach should therefore give results which are applicable to general human-human dialogue, but has some drawbacks. Firstly, no results concerning intonation can be obtained, although (as will be noted below) this might be important for disambiguating certain readings and forms of CR. Secondly, the absence of gesture information must mean some non-verbal interaction is missed – this may be important when examining answers to CRs as will be noted in section 3.4. Thirdly, while it is hoped that the spread of domains and speakers means that the results here can be taken to be generally applicable, the mostly non-domain-specific and entirely human-human nature of the dialogue means that care must be taken when extrapolating the results to human-computer dialogue systems, and especially those that deal with highly domain-specific or task-oriented dialogues. In particular domains, particular classes of words and phrases may be more important and thus more likely to be clarified (and their clarification more likely to be answered) than in the general conversation examined here.

Markup Scheme

The corpus was marked up according to the BNC's SGML conventions, with new tags inserted into the BNC files manually. 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. These possible values were initially taken to be those given by G&C's analysis, but evolved during the markup process as new CR mechanisms were identified; the final scheme consisted of the values described below in sections 3.2.2 and 3.2.3, together with an extra catch-all category *other* to deal with any otherwise uncategorisable phenomena.

The *source* attribute could take any numerical value and was set to the number of the sentence that was being clarified (according to the BNC sentence-numbering scheme).

Markup Details

Details of the markup tag syntax are shown below, together with brief examples of the form and reading classes. More detail of these classes, with examples from the corpus, are given below in sections 3.2.2 and 3.2.3.

Attribute		Possible Values	Example
rform	non	Non-Reprise	A:*"Did Bo leave?"* B:*"What did you say?"*
	lit	Literal Reprise	A:*"Did Bo leave?"* B:*"Did BO leave?"*
	sub	*WH*-Substituted Reprise	A:*"Did Bo leave?"* B:*"Did WHO leave?"*
	slu	Reprise Sluice	A:*"Did Bo leave?"* B:*"Who?"*
	frg	Reprise Fragment	A:*"Did Bo leave?"* B:*"Bo?"*
	gap	Reprise Gap	A:*"Did Bo leave?"* B:*"Did Bo . . . ?"*
	fil	Gap Filler	A:*"Did Bo . . . "* B:*". . . leave?"*
	oth	Other	
rread	cla	Clausal	*"Is it **Bo**$_i$ you're asking if i left?"*
	con	Constituent	*"Who do you mean by 'Bo'?"*
	lex	Lexical	*"Did you say 'Bo'?"*
	cor	Correction	*"Did you mean to say 'Mo'?"*
	oth	Other	
rsource	−	(any sentence number)	

Table 3.1: Clarification Request Markup Scheme

Listings 3.1 and 3.2 show an example of a marked-up CR in the SGML format used in the corpus, and the relevant part of the SGML Document Type Definition. The corpus marked up in this format has been made available to BNC license-holders.

```
<u who=PS1BY><s n="363">
<w PNP>I<w VBB>'m <w VVG>opening <w DPS>my <w DT0>own <w NN1>business
<w AV0>so <w PNP>I <w VVB>need <w AT0>a <w NN1>lot <w PRF>of <w NN1>money
</u>
<u who=PS1K6><s n="364" rform="slu" rread="lex" rsource="363">
<w NN1-VVG>Opening <w DTQ>what<c PUN>?
</u>
```

Listing 3.1: Example CR (example (54)) after markup

```
<!ATTLIST s
        id ID #IMPLIED
        n CDATA #IMPLIED
        p (Y | N) "N"
        rform (non | lit | sub | slu | frg | gap | fil | wot | oth) #IMPLIED
        rread (cla | con | lex | cor | oth) #IMPLIED
        rsource CDATA #IMPLIED
        TEIform CDATA "s" >
```

Listing 3.2: Excerpt from updated BNC Document Type Definition

Decision Process

Following the methods described by 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 are available for determination of CR source, for classification of form and for classification of reading: they are shown here in figures 3.1, 3.2 and 3.3 respectively.
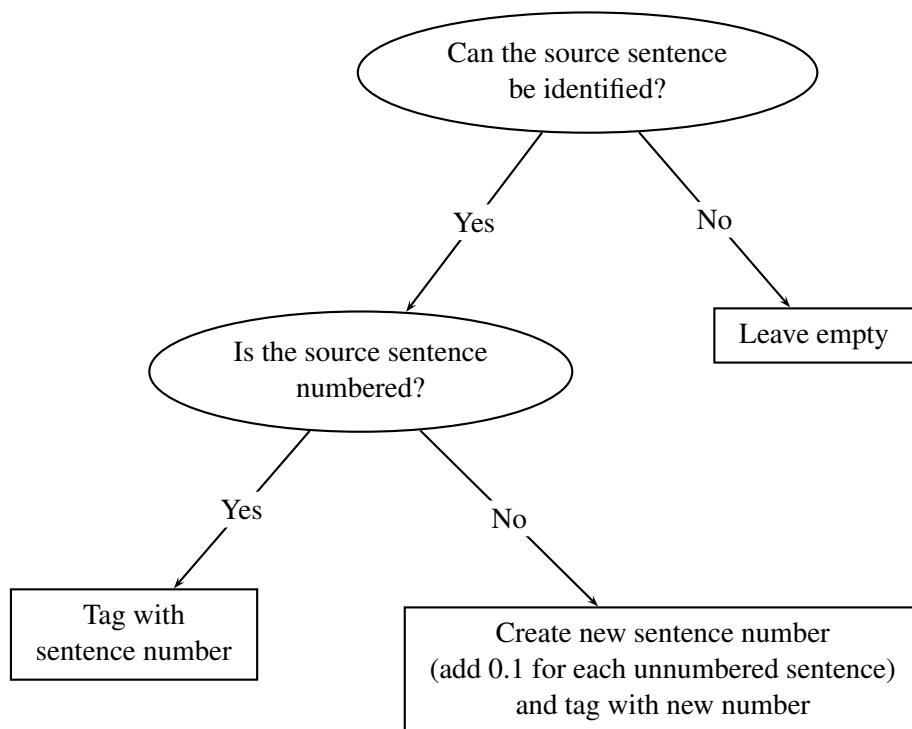
Figure 3.1: Decision Tree: CR Source

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

Tag as
`sub`

Tag as
`lit`

Yes

No

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

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

Yes

No

Tag as
`gap`

Yes

Tag as
`slu`

Tag as
`frg`

No

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

Yes

No

Tag as
`fil`

Tag as
`oth`

Figure 3.2: Decision Tree: CR Form

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

Yes   No

Tag as
`cla`

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

Yes   No

Tag as
`lex`

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   No

Tag as
`cor`

Tag as
`oth`

Figure 3.3: Decision Tree: CR Reading

Ambiguity of Reading   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 subsequent reaction of the speaker originally making the request (the CR *initiator*) was then used to judge whether this interpretation was correct (acceptable to the CR initiator). If the initiator gave no reaction to the contrary, the reading was assumed to have been acceptable. The following example (43) shows a case where the other participant's initial (clausal) interpretation was incorrect (the initiator is not satisfied), as a constituent reading was required. In such cases, both CRs were marked as constituent.

(43)[3]

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

(44)[4]

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

To ensure that this process is used correctly, 10 turns before and after the sentence being tagged were examined before the tagging decision was made. 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 displayed to the marker (but not themselves marked up).

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

The BNC sentence numbering scheme does not assign numbers to sentences containing no transcribed words. Such sentences are common where recording quality was poor or the environment was noisy – these sentences are marked in the BNC as `<unclear>` and given no number. Of course, these sentences are often unclear to other conversational participants, and therefore often cause CRs (usually with a lexical reading). In these cases, sentence numbers were assigned during tagging. Non-integer numbers were used, with values chosen to be consistent with the BNC numbering of surrounding sentences. For example, in example (45), the unclear sentence was given the number 589.1, and the source of the CR in sentence 590

---

[3]BNC file H5G, sentences 193–196
[4]BNC file KNV, sentences 548–551

was tagged with this number.

$(45)^5$ Peter: `<589>`   But he couldn't work out why I was in school?
            Muhammad: `<unclear>`
            Peter: `<590>`   **What?**

Reliability

The markup process was repeated (after an interval of several months) by the same annotator (myself), and was also performed by a naive annotator.[6] Both new versions were then compared to the original version, both in terms of raw agreement and the kappa statistic (see Carletta, 1996). The kappa statistic gives an indication of the level of agreement above that level which would be expected randomly: $\kappa = 100\%$ corresponds to perfect agreement, while $\kappa = 0$ corresponds to exactly that level of agreement which would be expected by chance.

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 3.2 below: CR source figures can be seen to be good (above 80% for both expert and naive annotators); figures for form & reading are less good, but are all above or close to the 80% level for both annotators. These levels are therefore not ideal, but probably good enough for the rather general conclusions we draw about form & reading distribution here.

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

Table 3.2: Markup Reliability

Examination of confusion matrices shows 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 3.2.2 below). Confusion of reading appears to be almost entirely due to genuine ambiguity in the corpus as presented.

It seems likely that much of the genuine ambiguity could be resolved by use of an audio corpus (or text corpus containing intonational markup) – intonation is useful when distinguishing the gap from the fragment form, and the clausal from the constituent reading. As far as the classification uncertainty goes, an improved classification scheme and instructions might help, in particular by conflating certain categories (we will return to this below).

---

[5] BNC file KPT, sentences 589–590
[6] Thanks are due to Charles Yee for his hard work and naivety.

3.2.2 Clarification Forms

The following forms were identified as possible means for CRs. The list may not be absolutely exhaustive, but gives good coverage of the CRs encountered in this corpus. This section lists the forms identified, and illustrates them with examples taken from the corpus.

Non-Reprise Clarifications (`non`)

Unsurprisingly, speakers have recourse to a non-reprise[7] form of clarification. In this form, the nature of the information being requested by the CR initiator is spelt out explicitly 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 (46) and (47).

(46)[8]

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.

(47)[9]

Leon:	Erm, your orgy is a food orgy.
Unknown:	**What did you say?**
Leon:	Your type of orgy is a food orgy.

Reprise Sentences (`lit`)

Speakers can form a CR by echoing or repeating a previous utterance in full, as shown in examples (48) and (49). This form corresponds to G&S's *reprise interrogative*.

(48)[10]

Orgady:	I spoke to him on Wednesday, I phoned him.
Obina:	**You phoned him?**
Orgady:	Phoned him.

(49)[11]

Gary:	No yo- `<pause>` no I'm getting paid for it.
Jake:	Eh?
	You get paid for it?
Lilias:	**You get paid**
Gary:	Aha
Lilias:	**for it?**
Gary:	Aye, I get a twenty five pound voucher next week for Marks and Spencers.

As already suggested by G&S, these repeats need not be verbatim. For one thing, index-

---

[7]Note that a *non-reprise* sentence need not necessarily be *non-elliptical*.

[8]BNC file KP4, sentences 521–524

[9]BNC file KPL, sentences 524–526

[10]BNC file KPW, sentences 463–465

[11]BNC file KPD, sentences 622–628

icals often change, as seen in both examples (48) and (49) above. Other changes may occur due to the use of phenomena such as VP ellipsis or anaphora, as shown in example (50).

(50)[12]

Anon 5:	Oh he's started this other job
Margaret:	**Oh he's started it?**
Anon 5:	Well, he he `<pause>` he works like the clappers he does!

Note that in these non-verbatim repeats, the meaning or reference of the changed phrases is preserved (or at least intended to be preserved – mistakes can be made when the hearer has not correctly understood the original source utterance).

WH-Substituted Reprise Sentences (`sub`)

A similar form is available where the sentence is repeated in full with the element under question replaced by a *wh*-phrase, as illustrated by examples (51) and (52).

(51)[13]

Unknown:	He's anal retentive, that's what it is.
Kath:	**He's what?**
Unknown:	Anal retentive.

(52)[14]

Blake:	Everybody at school says mozzarella cheese is disgusting `<pause>` I take no notice of them
Skonev:	Well I should think that's quite right.
	They probably, they probably get it given to you on your school pizzas `<unclear>`
Antony:	**What cheese is disgusting?**
Skonev:	Mozzarella, the one you're eating

Again, the repeated part need not be verbatim, but reprises the intended meaning of the original utterance.

Reprise Sluices (`slu`)

This form is an elliptical *wh*-construction as already introduced in chapter 2 (and described by G&S), in which a bare *wh*-phrase is used to reprise a particular phrase in the source utter-

---

[12]BNC file KST, sentences 455–457
[13]BNC file KPH, sentences 412–414
[14]BNC file KR1, sentences 470–474

ance.[15]

(53)[16]
Sarah:	Leon, Leon, sorry she's taken.
Leon:	**Who?**
Sarah:	Cath Long, she's spoken for.

(54)[17]
Sheila:	No he's, he's being moved to troop fifteen
Wendy:	**To where?**
Sheila:	Troop fifteen
Wendy:	Oh

There may be a continuum of forms between *wh-substituted reprise sentences* and *reprise sluices*. Consider the following exchange (example (55)):

(55)[18]
| 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, but such decisions are not easy and were the cause of some of the markup disagreements noted in section 3.2.1 above. Conflating these forms might remove this source of disagreement, but will depend on whether syntactic and semantic analyses are compatible – this is discussed in section 3.2.5 below.[19]

Reprise Fragments (`frg`)

This elliptical bare fragment form corresponds to that described as *elliptical literal reprise* by G&S and *clarification ellipsis* by G&C: a bare fragment is used to reprise a particular phrase in the source utterance.

(56)[20]
Lara:	There's only two people in the class.
Matthew:	**Two people?**
Unknown:	For cookery, yeah.

---

[15]As already mentioned in chapter 2, this reprise sluice class of CRs does not include *direct* sluices, which are not taken to be CRs – rather than querying intended form or content, they request further information about an existentially quantified expression. See section 2.5.2.

[16]BNC file KPL, sentences 347–349

[17]BNC file KR0, sentences 442–445

[18]BNC file KSV, sentences 363–364

[19]A similar continuum might be present between literal reprises and reprise fragments.

[20]BNC file KPP, sentences 352–354

(57)[21]	Catriona:	God I hope I don't look like big Kath `<unclear>` blessing if you did.
	Jess:	**Blessing?**
	Catriona:	Mm.
	Jess:	What you would like to look like her?

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

(58)[22]	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 (58), *"Emma?"*), they were not distinguished from fragments in the classification scheme. In terms of analysis, such sentences seem best treated as a reprise sluice followed by a reprise fragment, rather than as a separate form.

Reprise Gaps (`gap`)

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 (59).

(59)[23]	Laura:	Can I have some toast please?
	Jan:	**Some?**
	Laura:	Toast

Initially this may seem to resemble the reprise fragment, but it has been classified as a separate form nonetheless. Firstly, it does represent a different method for clarifying a particular phrase: reprising whatever immediately precedes it, rather than the phrase itself. In example (59), the word intended to be clarified is *toast*; a reprise fragment CR would involve reprising that word (e.g. *"Toast?"*), whereas a reprise gap CR involves reprising the previous word (*"Some?"*).

Secondly, personal judgements suggest that this form is intonationally distinct from the reprise fragment form. As there is no intonational information in the BNC, this cannot be verified from this study, but the gap form appears to be necessarily associated with a high flat "continuation" tone (with no final rise). As might be expected given this distinction, no misunderstandings of gap-CRs were discovered during our corpus analysis – although as only two examples were found, this fact cannot be regarded as significant evidence in itself.

---

[21]BNC file KP6, sentences 494–497
[22]BNC file KSW, sentences 698–700
[23]BNC file KD7, sentences 392–394

Gap Fillers (`fil`)

The *filler* form is used by a speaker to ask about or suggest material which might 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 (60)) or because the CR initiator interrupts.

(60)[24]

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

(61)[25]

TF:	I'm pretty sure that the
D:	**Programmed visits?**
TF:	Programmed visits, yes, I think they'll have been debt inspections

This form is therefore slightly different from the others in this classification, in that it does not clarify a part of the original utterance as actually presented, but instead a part that was originally intended by the speaker but not produced. As this still fits with the spirit of G&S's analysis of reprise sentences as querying intended content, it has been included here.

Conventional (`wot`)

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?"*:

(62)[26]

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

(63)[27]

Leslie:	`<clears throat>` `<pause>` I didn't know it was that high.
Steve:	**What?**
Leslie:	I wouldn't rate it that high.

(64)[28]

Unknown:	You're making it up sir `<unclear>` story.
Richardson:	**Pardon?**
Unknown:	You making that up?
Richardson:	No.

---

[24]BNC file KPJ, sentences 555–559
[25]BNC file KS1, sentences 789–791
[26]BNC file KPK, sentences 580–582
[27]BNC file KSU, sentences 447–449
[28]BNC file KP3, sentences 891–894

3.2.3 Clarification Readings

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

Clausal (`cla`)

The *clausal* reading takes as the basis for its content the *content of the conversational move* made by the utterance being clarified: asking a question, asserting a proposition etc.

This reading is paraphrasable roughly as *"Are you asking/asserting P?"*, *"Is it X about which you are asking/asserting P(X)?"*, or *"For which X are you asking/asserting P(X)?"* (depending on whether the question being asked is a yes/no or *wh*-question). 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. Examples of literal reprise sentence, fragment and sluice are shown here with imagined paraphrases which illustrate the clausal nature of the intended reading.

(65)[29]

Orgady:	I spoke to him on Wednesday, I phoned him.
Obina:	**You phoned him?**
Orgady:	Phoned him.
$\rightsquigarrow$	*"Are you asserting that **you phoned him**?"*

(66)[30]

Lara:	There's only two people in the class.
Matthew:	**Two people?**
Unknown:	For cookery, yeah.
$\rightsquigarrow$	*"Is it **two people** you are asserting are in the class?"*

(67)[31]

Sarah:	Leon, Leon, sorry she's taken.
Leon:	**Who?**
Sarah:	Cath Long, she's spoken for.
$\rightsquigarrow$	*"**Who** is it you are asserting is taken?"*

Constituent (`con`)

The other possible reading given an analysis by G&C is a *constituent* reading whereby the content of *a constituent* of the previous utterance is being clarified.

---

[29]BNC file KPW, sentences 463–465
[30]BNC file KPP, sentences 352–354
[31]BNC file KPL, sentences 347–349

This reading corresponds roughly to *"What/who is 'X'?"*, *"What/who do you mean by 'X'?"*, or *"Is it Y that you mean by 'X'?"*, a question about the intended semantic content or reference of a (sub-)utterance.

(68)[32]

	Frances:	She likes boys called Leigh, named Leigh, Leigh `[name]`, `[name]`, Leigh `[name]` `<pause>` Bill Leigh `[name]`, B J.
	Ben:	**B J.**
	Frances:	She, she's writing a note
	Ben:	**B J?**
	Frances:	you know Ash, B J
	Ben:	What?
	Frances:	B J.
	Ben:	Don't mean nothing.
	Frances:	You know B J, it stands for blow job right.
	$\leadsto$	*"What do you mean by 'BJ'?"*
	$\leadsto$	*"What is a 'BJ'?"*

(69)[33]

	Rupert:	Oh no!
	Jimmy:	**What?** `<pause>`
	Rupert:	Something's going on.
	$\leadsto$	*"What do you mean by that?"*

Lexical (`lex`)

Another possibility appears to be a *lexical* reading. This seems 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.[34]

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/segment in the source utterance, rather than a part of the semantic content of the utterance.

(70)[35]

	Anon 6:	here that Sassafras has been `<pause>` named potentially unsafe for consumption. So, don't put any in your mouth.
	Margaret:	**Saxa-what**?
	Anon 5:	Saxa frall `[sic]` that's a plant!
	$\leadsto$	*"What X did you utter 'Saxa-X'?"*

---

[32]BNC file KSW, sentences 611-619

[33]BNC file KP0, sentences 439–441

[34]In fact, a more suitable name for this reading might be *form identification*. *Lexical* will be used here to maintain continuity, but this name should not be taken to imply that only single words can be asked about.

[35]BNC file KST, sentences 499–502

	Cassie:	Give me this a minute. `<playing music dur=15>`.
	Bonnie:	Is your mum gonna hear that?
	Cassie:	**What?**
	Bonnie:	`<unclear>`
(71)[36]	Cassie:	**Do I let my mum hear it?**
	Bonnie:	Yeah.
	$\rightsquigarrow$	*"What did you say?"*
	$\rightsquigarrow$	*"Did you say 'do I let my mum hear it'?"*

Corrections (`cor`)

The correction reading is paraphrasable as *"Did you intend to/should you have uttered X (instead of Y)?"*. This is therefore similar to the lexical reading in that it queries surface form rather than semantic content, but is distinguished by the fact that it queries a possible replacement or substitution of one part of the original form with another.

	Anon 3:	Last year I was fifteen for the third time round.
	Grace:	Yeah.
		`<laugh>` Fifteen for the first time round.
(72)[37]	Anon 3:	**Third.**
	Grace:	Third time round.
	Anon 3:	Third time round.

	Shelley:	My `<pause>` er, that's it, my sister's
	Unknown:	Why are you writing problems?
(73)[38]	Shelley:	boyfriend said I'm a common cow and have a got a big nose.
	Unknown:	Did she?
		Did he?
	Shelley:	Yeah.

	Frances:	You know Amy?
	Ben:	Yeah.
(74)[39]	Frances:	Do you reckon that er is, her sister?
		Her brother I mean?
	Ben:	Amy?
		`<unclear>` Mm.

The lexical nature of this reading as defined above is due to the fact that only corrections that seemed to fit with this lexical paraphrase were found in this corpus of examples. However,

---

[36]BNC file KP4, sentences 353–357
[37]BNC file KPE, Sentences 326–331
[38]BNC file KPG, sentences 485–490
[39]BNC file KSW, sentences 528–533

it seems quite possible that corrections can in fact have clausal or constituent sub-type too: paraphrases such as *"Did you intend to assert P(X) (instead of P(Y))?"* and *"Did you intend to refer to X (instead of Y)?"* seem perfectly plausible as CRs. This reading might therefore be better described not as a separate reading but as a particular usage of those already established – see the suggested analysis below.

3.2.4 Results

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

Form/Reading: The distribution of CRs by form and reading is shown in full as raw counts in table 3.3 for all domains. The distributions are also presented as percentages of all CRs found in table 3.4 (all dialogue domains) and table 3.5 (demographic only). 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 of domains, indicating that corpus size is large enough to give repeatable results.

CSS Distance Separation between CR and source 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 might be expected; less obviously, 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 3.6 and 3.7, as percentages in figures 3.4 and 3.5, and as cumulative percentages in figures 3.6 and 3.7.

	non	lit	sub	slu	frg	gap	fil	wot	oth	*Total*
cla	11	26	6	48	104	0	0	0	2	197
con	32	0	0	0	7	0	0	21	0	60
lex	3	0	9	6	1	2	17	107	0	145
cor	3	2	0	0	5	0	0	0	0	10
oth	0	0	0	0	4	0	0	2	0	6
Total	49	28	15	54	121	2	17	130	2	418

Table 3.3: CR form vs. type – all domains

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,

	non	lit	sub	slu	frg	gap	fil	wot	oth	*Total*
cla	2.6	6.2	1.4	11.5	24.9	0	0	0	0.5	47.1
con	7.7	0	0	0	1.7	0	0	5.0	0	14.4
lex	0.7	0	2.2	1.4	0.2	0.5	4.1	25.6	0	34.7
cor	0.7	0.5	0	0	1.2	0	0	0	0	2.4
oth	0	0	0	0	1.0	0	0	0.5	0	1.5
Total	11.7	6.7	3.6	12.9	29.0	0.5	4.1	31.1	0.5	(100)

Table 3.4: CR form and type as percentage of CRs – all domains

	non	lit	sub	slu	frg	gap	fil	wot	oth	*Total*
cla	2.6	6.0	1.6	12.2	24.6	0	0	0	0.5	47.5
con	6.0	0	0	0	1.8	0	0	5.4	0	13.2
lex	0.8	0	2.1	1.6	0.3	0.5	3.4	26.9	0	35.6
cor	0.8	0.5	0	0	1.3	0	0	0	0	2.6
oth	0	0	0	0	0.8	0	0	0.5	0	1.3
Total	10.2	6.5	3.7	13.8	28.8	0.5	3.4	32.8	0.5	(100)

Table 3.5: CR form and type as percentage of CRs – demographic portion

giving support to the claim that processing of CRs is important for a dialogue system.[40]

The most common forms of CR can be seen to be the conventional and reprise fragment forms, with each making up around 30% of CRs. Non-reprise CRs and reprise sluices are also common, each contributing over 10% of CRs. Other forms (*wh*-substituted and reprise sentences, fillers and gaps) 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. Corrections are much less common (2–3%).

This initially suggests that an automated dialogue system which can deal with fragments, sluices and reprise sentences (G&C's analyses described in section 2.3.4), together with conventional and non-reprise CRs, could give reasonable coverage of expected forms; and that covering lexical readings (in addition to the clausal and constituent analyses already given) will be required, at least for some forms.

Coverage

The coverage of the corpus by the forms and readings listed in sections 3.2.2 and 3.2.3 is good, with only 0.5% of CR forms (2 sentences) and about 1.5% of CR readings (6 sentences) being classified as *other*.

---

[40]Although this proportion is calculated on the basis of human-human dialogue, David Traum (p.c.) has indicated that a similar proportion was observed during the TRAINS experiments (see e.g. Heeman and Allen, 1995).

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 (example (75)) or conventional form (example (76)).

(75)[41]

Sarah:	what did you get in the table assess yesterday?
Marsha:	table assess twenty eight, erm twenty eight, eighteen
Carla:	**twenty eight** `<laugh>`
Sarah:	eighteen

⤳   ?*"Are you telling me you got twenty-eight?"*
⤳   *"It's funny/ridiculous to say you got twenty-eight"*

(76)[42]

Muhammad:	Go up to Miss thingy.
Peter:	**What?**
Muhammad:	Go on.
	Go on. `<laugh>`
Peter:	Who Miss `[name]`?

⤳   ?*"What did you say?"*
⤳   *"How can you suggest something so naughty?"*

Intuitively it seems that these readings can be treated as standard readings (as shown in the first suggested paraphrase in each case, for example (75) a clausal reading and for example (76) a lexical reading) which are then enriched with a further level of illocutionary force given by use in context and pragmatic inference to give the full speaker's meaning (as shown in the second suggested paraphrase in each case). This idea is perhaps reinforced by imagining possible responses to the CRs: in example (75), an answer *"Yes"* could certainly be interpreted as answering the standard question and meaning something like *"Yes, I am telling you I got twenty-eight"*. This suggests that these standard readings are available, and a reasonable analysis might be to assign them as basic and allow for further inference if necessary.

Jens Allwood (p.c.) points out that several further readings are possible (for example, a "courtroom" reading which is not really intended to clarify but to confirm strongly, particularly for the benefit of other hearers), although no examples were observed within the scope of this corpus. Again, a reasonable approach might be to treat them as standard (usually clausal) readings, with further inferential effects.

Of the 2 sentences left unclassified for form, one appears to be an unusual conventional form example (77), and one an interesting example of a literal reprise of an unuttered but

---

[41]BNC file KP2, sentences 376–379
[42]BNC file KPT, sentences 506–510

implied sentence example (78):

(77)[43]

Lara:	. . . next Tuesday at five o'clock there's a lowdown.
Matthew:	Eh?
Lara:	With you on it.
Matthew:	With me?
Lara:	Yeah the model of you.
Matthew:	**How?**
Lara:	How what?
Matthew:	What model of me?
Unknown:	What the Thinker?
Lara:	The Thinker, that's the lowdown isn't it?
Matthew:	No.

(78)[44]

Anon 2:	You haven't given me the bits missing from the Mirror.
Anon 3:	Well that's erm `<unclear>`
Anon 2:	Well I'm sorry but my Daily Mirror was dev delivered this morning without the television supplement and without the comic and I want them
Anon 3:	the actual page?
Anon 2:	The whole supplement that come on Saturdays
Anon 3:	`<unclear>`
Anon 2:	That part
Anon 3:	**That part wasn't in?**
Anon 2:	This part was not in. And there's usually a free comic as well, and that wasn't in either.

Example (77) could be treated by analysing (via a suitable lexicon) as a conventional CR (if that is indeed what it is – it is perhaps significant that the other participant in the dialogue also seems to have trouble interpreting this example). Example (78) could be treated as a literal reprise sentence, but only by a system which was capable of relating *"X was delivered without Y"* to *"Y wasn't in X"*. Alternatively it could perhaps be treated as not a CR at all, but a direct question, which in this case seems plausible: *"Was that part not in?"*, rather than *"Are you telling me that part wasn't in?"*.

In general, though, the small number of these problematic cases is encouraging: it seems that the ontology of readings and forms given here is sufficient to cover the large majority of examples.

Form/Reading Correlation

The non-reprise form appears to be able to carry any reading – of course, with this form the reading is spelt out directly, so we are not so concerned with examining trends. However, it does seem that lexical problems (which must be common, given the high number of lexical

---

[43]BNC file KPP, sentences 442-452
[44]BNC file KNS, sentences 487–495

CRs overall) are unlikely to be clarified this way: the lexical reading is uncommon for this form (only 6% of non-reprise CRs are lexical, as opposed to 35% overall), and statistically significantly so: comparing the distribution of lexical vs. non-lexical readings between non-reprise CRs and all other CRs with a $\chi^2_{(1)}$ test gives $p < 0.1\%$[45]. The apparently low number of correction readings is actually relatively high given the overall rarity of this reading, but not significantly so ($\chi^2_{(1)}$ gives $p = 7\%$).

In contrast, the conventional form is most likely to have a lexical reading, and the clausal reading seems impossible (unsurprisingly, given the definition that the clausal reading involves the conversational move made, and the fact that the conventional form seems to signal complete lack of understanding). The high likelihood of lexical vs. constituent goes against the general trend: 82% of conventional CRs are lexical and only 16% constituent, whereas the equivalent figures over all other readings are a much more even 13% and 14%. This is statistically significant ($\chi^2_{(1)}$ gives $p < 0.1\%$).

The *wh*-forms, sluices and *wh*-substituted reprise sentences, appear always to be satisfactorily interpretable by a clausal or lexical reading. Sluices strongly prefer clausal readings to lexical (89% of cases), and we can have confidence in this: $\chi^2_{(1)}$ gives $p < 0.1\%$. *WH*-substituted sentences seem to prefer lexical readings, although the effect is less strong (60%) and not statistically significant: $p = 16\%$.

Literal reprise sentences seem to greatly prefer clausal readings (neither constituent nor lexical were seen), and both preferences are significant: $\chi^2_{(1)}$ gives $p = 0.3\%$ when comparing the clausal/constituent distribution with all other CR forms, with $p < 0.1\%$ for the clausal/lexical bias. Similarly, comparing the distribution of all three readings together, with a $\chi^2_{(2)}$ test, gives $p < 0.1\%$.

Reprise fragments also strongly prefer clausal readings (86% of cases) over constituent and lexical. Again this is significant ($p < 0.1\%$ for both preferences individually and across all three), but this time the preference is not quite as strong: both constituent and lexical readings are possible, although unlikely (6% and 1% of fragments respectively).

Gap and filler forms both appear only to be used with a lexical reading, but the low number of examples of these forms observed means it may be dangerous to attempt to draw any firm conclusions. The absence of both clausal and constituent readings seems significant for the filler form ($p < 1\%$ for both individually, $p < 0.1\%$ across all three readings), but not so for the gap ($\chi^2_{(2)} = 17\%$).[46]

Note also that the constituent reading only occurs with the non-reprise, conventional and reprise fragment forms. Even then, it is much less common than the clausal or lexical read-

---

[45]The χ^2 values in this section give the probability that the reading distribution is *independent* of the form – i.e. that the reading distribution for the particular form in question is merely a reflection of the overall reading distribution. A criterion of $< 5\%$ is usually taken as a reasonable indication of significant dependence.

[46]Intuitively, it does seem that other gap readings, possibly querying semantic content in some way, may be possible. In the following invented exchange, the question is asked with the typical gap intonation, and asks a question about what might come after the echoed word, but cannot be a simple question about which word was

ings.

In summary, it appears that many readings are available for some forms (for example, the reprise fragment form, which appears to allow all readings), so disambiguation between readings will be important for a dialogue system. Fortunately, some significant correlations between form and reading can be seen which will help with this. In particular:

- Literal reprises only take clausal readings.

- Fillers only take lexical readings.

- *WH*-forms have only clausal or lexical readings.

- Fragments and sluices usually take clausal readings.

- Conventional CRs usually have lexical readings.

Whether we can take these as general conclusions which might hold across different domains (and perhaps even different languages) is a slightly different matter. In the case of the *wh*-forms, perhaps we can: the nature of the *wh*-questions asked by these forms seems to make the clausal and constituent readings indistinguishable, so that only one might be required: given a source utterance *"I want to go to Paris"* and a reprise sluice *"Where?"*, while the readings *"Where do you mean by 'Paris'?"* and *"Where is it you're asserting you want to go to?"* might be formally different, they seem to ask for (and be answerable by) exactly the same information. The fact that all have been classed as clausal here may be an effect of the ordering of the markup decision tree, but this reading does seem the more natural, at least for the full *wh*-substituted sentence form – the clausal paraphrase above seems much more natural for a reprise *"You want to go to WHERE?"* than the constituent paraphrase does. For fillers, too, perhaps we can generalise: they seem only to make intuitive sense as asking a question about lexical identity, so perhaps we can take this as a general principle. It also seems impossible that conventional CRs, given their general utterance-level nature, could ask a clausal type of question.

More care should probably be taken with the others, though. As will be discussed in section 4.1.1, there may be a possibility of literal reprises asking constituent-type questions; also, given that reprise fragments certainly do take constituent readings, and that there may be a continuum of forms between fragment and full sentence, entirely ruling out constituent readings for the full sentence form might be problematic. There also seems no reason to assume that the relative likelihood of the fragment readings will be domain-independent.

---

uttered next:

	A:	I saw Bo yesterday.
(79)	B:	**Bo . . . ?**
	A:	Bo Smith.

Given the lack of evidence, we will not devote time to analysing such examples, but as noted in section 3.2.5 below, they would be treatable given a contextual coercion operation which produces a suitable question.

CR-Source Separation Distance

The maximum CSS distance observed was 15 sentences (14 turns). However, only one example of this distance was observed, and one example of distance 13 sentences (11 turns) – otherwise all CSS distances were below 10 sentences (see table 3.6). The vast majority (about 80%) of CRs had a CSS distance of 1 sentence (i.e. were clarifying the immediately preceding sentence – see figures 3.4 and 3.6), and over 96% had a distance of 4 sentences or less. Similarly in turns, 84% had a distance of 1, and 98% a distance of 4 or less – see figures 3.5 and 3.7.

Only one example had a distance of 0 sentences (a self-correction within the same sentence). Zero distance is more common for turns, as one might expect: 11 examples (nearly 3% of the total) had a distance of 0 turns, and about half of these (6) were self-corrections. The rest were self-referential clarification which, while not strictly correcting a previous sentence, asked or asserted clarificatory information about words or reference:

$(80)^{47}$

	Eddie:	Did you, did tha- that come up about eating them?
		Poppy seeds?
	Anon 2:	Someone said about the poppy, you can eat poppy seed . . .
	↝	*"By 'them' I mean poppy seeds"*
	↝	*"It's **poppy seeds** $_X$ that I'm asking if it came up about eating X?"*

$(81)^{48}$

	Unknown:	Is it such as a good school as that?
		Are you such an open society.
		I mean do you, for example, gives girls a fair chance in your school?
	Unknown:	Yes, I think we actually positively discriminate to encourage girls . . .
	↝	*"By 'are you such an open society' I mean do you give . . . "*

---

[47]BNC file KPB, sentences 558–560
[48]BNC file KRG, sentences 1518–1521

Distance	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	-	Total
All domains	1	331	39	15	12	3	4	0	2	2	0	0	0	1	0	1	7	418
Demographic	0	304	37	15	10	3	4	0	2	2	0	0	0	1	0	1	7	386

Table 3.6: Number of CRs vs. CSS Distance (Sentences)

Distance	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	-	Total
All domains	11	347	18	18	9	2	1	1	1	0	1	1	0	0	1	0	7	418
Demographic	9	320	17	16	9	2	1	1	1	0	1	1	0	0	1	0	7	386

Table 3.7: Number of CRs vs. CSS Distance (Turns)
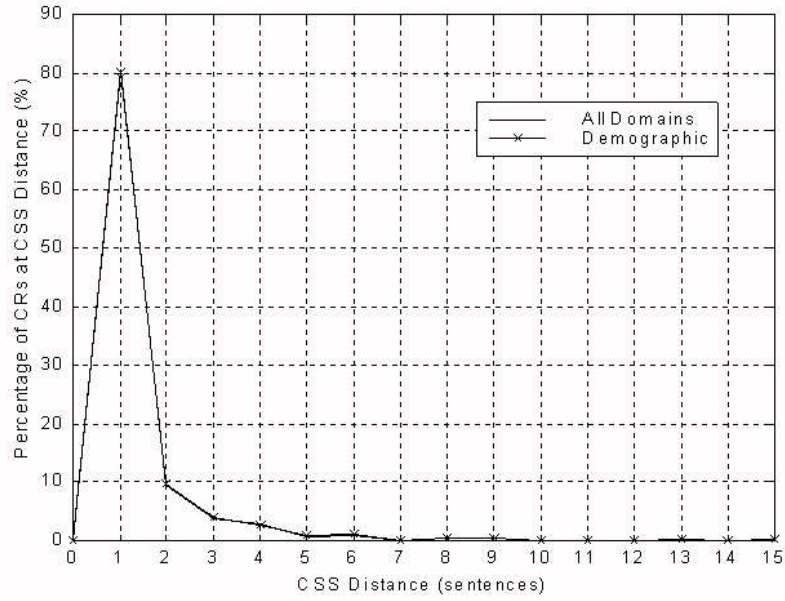
Figure 3.4: Percentage of CRs vs. CSS Distance (Sentences)
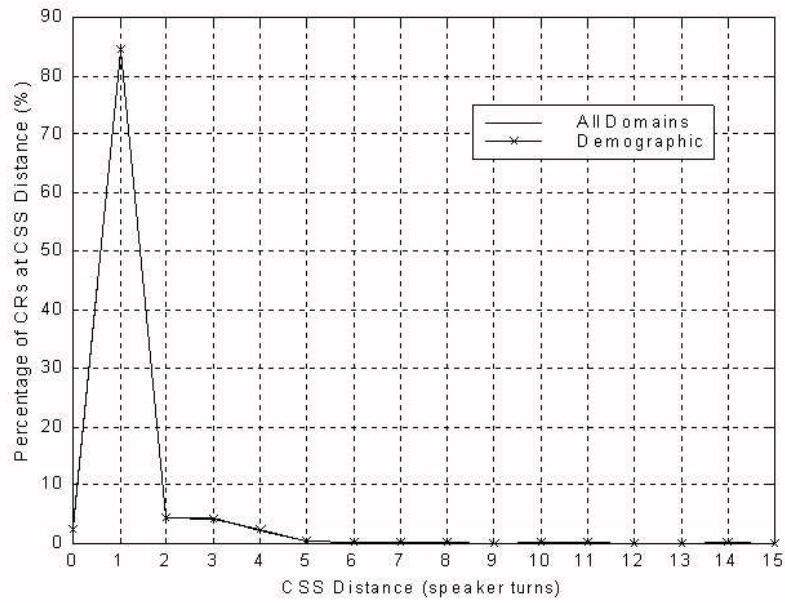
Figure 3.5: Percentage of CRs vs. CSS Distance (Turns)
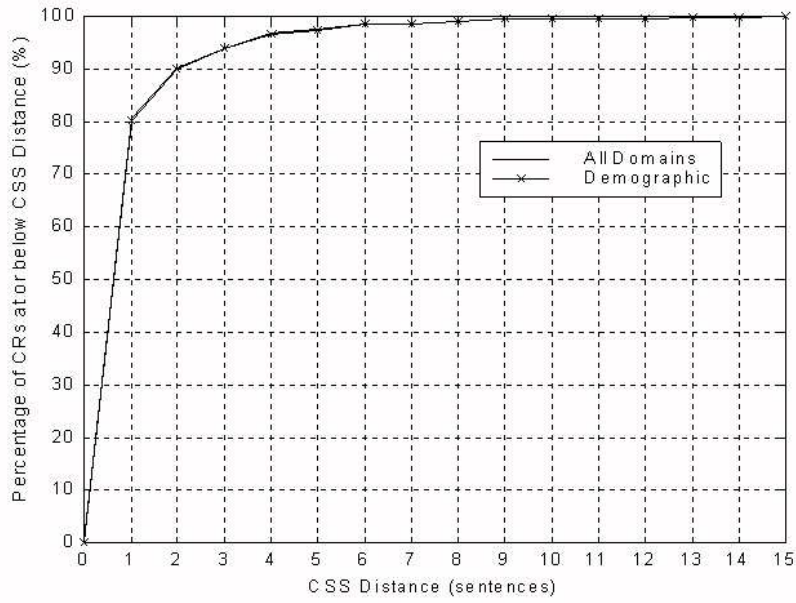
Figure 3.6: Cumulative Percentage of CRs vs. CSS Distance (Sentences)
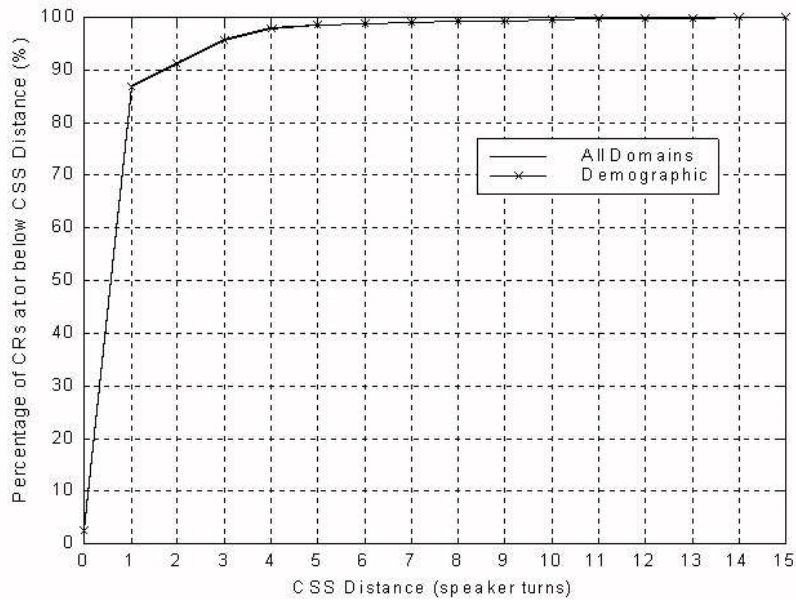
Figure 3.7: Cumulative Percentage of CRs vs. CSS Distance (Turns)

Another difference between the figures for sentences and turns is worth noting. While the number of occurrences decreases smoothly with CSS distance (above 1) for sentences (see figure 3.4), we can see from figure 3.5 that this is not quite true for turns: a distance of 3 seems as common as a distance of 2. Figure 3.8 shows this difference directly. This may be an indication that clarification of another participant's contributions is generally 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). It also suggests that a measure of CSS distance with more explanatory power might be one which takes into account discourse structure in some way (perhaps examining the number of intervening complete or incomplete adjacency pairs). However, as the basic findings here give enough information for the current purposes of determining likely memory requirements in a dialogue system, and as reliably annotating discourse structure on the BNC is a significant task in itself, this has not been attempted here.
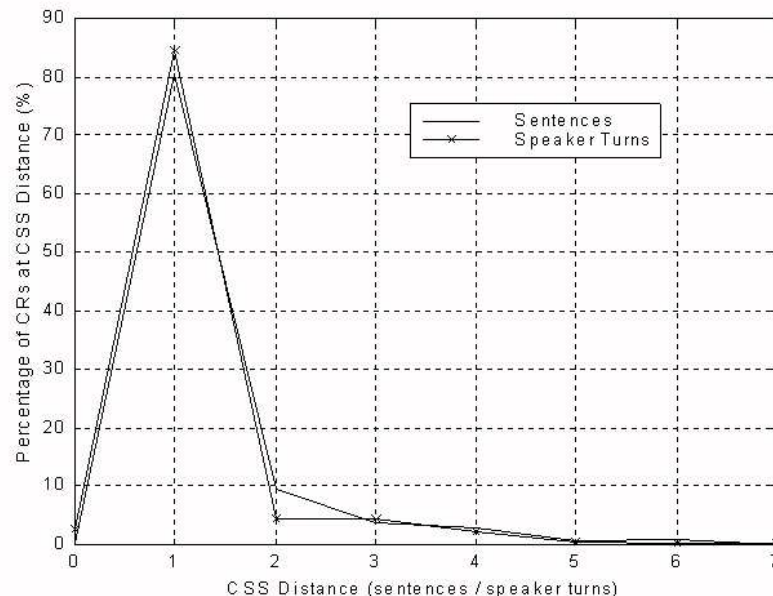
Figure 3.8: Percentage of CRs vs. CSS Distance (Sentences vs. Turns)

Number of Participants   It should be noted that the two long-distance cases (CSS > 10 sentences) were both seen in one dialogue which had several speakers present (the dialogue was in a classroom situation with many people talking and one speaker attempting to clarify an utterance by the teacher). These kind of dialogues are presumably not representative of the situation expected with an automated dialogue system and therefore only two participants; in particular one might expect CSS distances to be shorter in two-party situations. Results were

therefore divided into two-party and multi-party dialogues for comparison.

However, identifying the number of participants in the BNC dialogue transcripts is not entirely straightforward. Firstly, there are several tagging errors (e.g. cases where all utterances in what appears to be at least a two-party dialogue are tagged as the same speaker). Secondly, while each file is divided into several divisions (tagged `<div>`) which correspond to contiguous sections of recording (and thus roughly to separate scenarios or dialogues), the nature of most of the recordings means it is not possible to be sure if or when participants enter or leave. Thirdly, there is a problem with unknown speakers: while the transcription scheme marks each utterance with an identifier corresponding to the speaker, there are two dedicated identifiers used in cases where the transcriber could not determine the speaker (one for single speakers, one for multiple simultaneous speakers). There is then no way of telling whether utterances marked with these identifiers were produced by one of the other (already identified) participants, or by a further otherwise unidentified participant. To be conservative, these identifiers were considered as participants in their own right, and all participants in a contiguous division were counted; this means that some two-party dialogues will have been counted as multi-party, and that the number of two-party results becomes low, but does ensure that only genuine two-party dialogues were considered as such.

There is a further complication with possible silent participants (who might still be considered part of the dialogue even though not recorded as speaking in the recorded/transcribed section): simply counting the identified speakers in the transcription will not take these into account. The BNC header information does give a way of getting at this, as it lists participants for each division, as noted by the person making the recording. However, in several cases there appear to be errors in this information (e.g. participants marked as making utterances in the transcript body itself but not listed in the header). Results have therefore been given both as determined by the *header* and as determined by *direct* counting of speakers marked in the body of the transcripts.

The results are shown in tables 3.8 and 3.9 for sentences and speaker turns respectively. As expected, the maximum CSS distance for two-party dialogue is shorter than for multi-party dialogue, now being 6 sentences or 4-6 turns (depending on the method of determining number of participants). However, it does not appear that the effect is strong enough to limit CRs to the immediately subsequent sentence/turn (i.e. to a CSS distance of 1). In fact, as shown in figures 3.9 and 3.10, distances of 2-4 sentences/turns do not seem significantly rarer for two-party than for multi-party dialogues. In the two-party examples, while the majority are, again, at a CSS distance of 1 (72-74% in sentences, 83-84% in turns), it is not until a distance of 4 sentences/turns that 95% of CRs are included.

Distance	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	-	*Total*
≤ 2 *participants (header)*	0	21	5	1	1	0	1	0	0	0	0	0	0	0	0	0	0	29
> 2 *participants (header)*	1	310	34	14	11	3	3	0	2	2	0	0	0	1	0	1	7	389
≤ 2 *participants (direct)*	1	27	5	1	2	1	1	0	0	0	0	0	0	0	0	0	0	38
> 2 *participants (direct)*	0	304	34	14	10	2	3	0	2	2	0	0	0	1	0	1	7	380

Table 3.8: Number of Participants vs. CSS Distance (Sentences)

Distance	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	-	*Total*
≤ 2 *participants (header)*	1	23	2	1	1	0	1	0	0	0	0	0	0	0	0	0	0	29
> 2 *participants (header)*	10	324	16	17	8	2	0	1	1	0	1	1	0	0	1	0	7	389
≤ 2 *participants (direct)*	2	30	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	38
> 2 *participants (direct)*	9	317	16	16	7	2	1	1	1	0	1	1	0	0	1	0	7	380

Table 3.9: Number of Participants vs. CSS Distance (Turns)
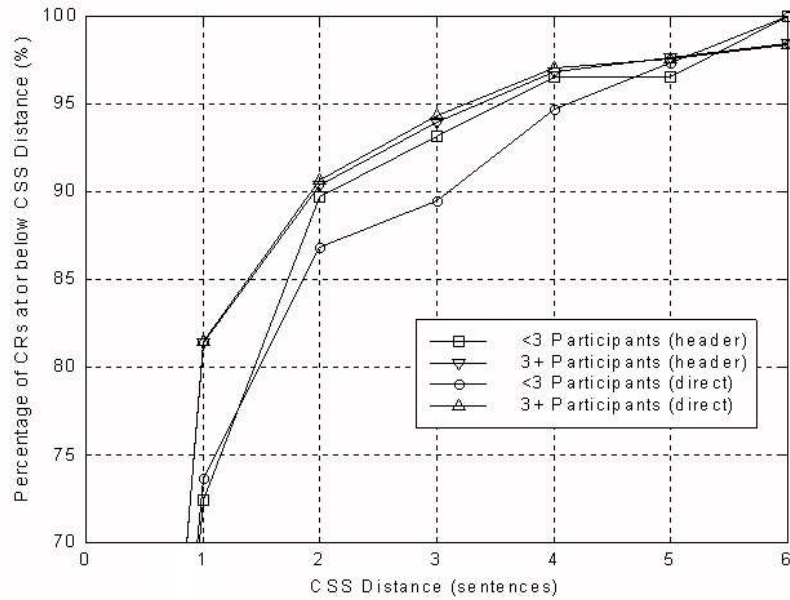
Figure 3.9: Percentage of CRs vs. CSS Distance / No. of Participants (Sentences)
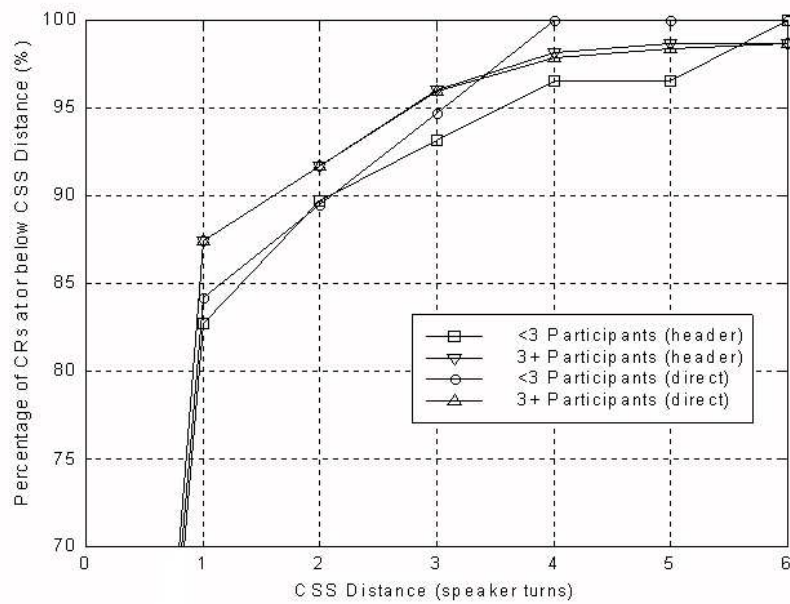
Figure 3.10: Percentage of CRs vs. CSS Distance / No. of Participants (Turns)

Form/CSS Correlation

There is a correlation between CSS distance and CR form: some forms are much more likely than others to be querying the immediately preceding turn. Figures (over all domains) for CSS distance per form are shown in table 3.10, and as percentages per form, excluding zero and unclassified distances, in table 3.11.

	0	1	2	3	4	>4	-	*Total*
non	5	32	5	3	0	2	2	49
lit	1	20	2	2	1	2	0	28
sub	0	13	2	0	0	0	0	15
slu	0	45	0	5	3	1	0	54
frg	5	100	4	7	3	2	0	121
gap	0	2	0	0	0	0	0	2
fil	0	16	0	1	0	0	0	17
wot	0	117	5	0	2	1	5	130
oth	0	2	0	0	0	0	0	2
Total	11	347	18	18	9	8	7	418

Table 3.10: CR Form vs. CSS Distance (turns)

	1	2	3	4	>4	*Total*
non	76.2	11.9	7.1	0	4.8	(100)
lit	74.1	7.4	7.4	3.7	7.4	(100)
sub	86.7	13.3	0	0	0	(100)
slu	83.3	0	9.3	5.6	1.9	(100)
frg	86.2	3.4	6.0	2.6	1.7	(100)
gap	100.0	0	0	0	0	(100)
fil	94.1	0	5.9	0	0	(100)
wot	93.6	4.0	0	1.6	0.8	(100)
oth	100.0	0	0	0	0	(100)

Table 3.11: CR Form vs. CSS Distance (turns) – percentages per form

We can see that in general, the more elliptical forms are more likely to have a low CSS distance: over 90% of gaps, fillers and conventional CRs query the immediately preceding turn. Reprise fragments and sluices are nearly as likely (83-86%) to have CSS distance 1, but then do not drop off quite as smoothly as the others. The forms most likely to have high CSS distances are the (less elliptical) non-reprise and the literal reprise form.

This is encouraging from the point of view of disambiguation of source: the non-reprise and literal reprise forms are more likely to spell out which turn is being queried (by echoing it for the literal reprise, and by the explicit nature of the non-reprise). For those that are more likely to be ambiguous (such as the gap, fragment, sluice and conventional forms), a

strategy of assuming the most recent turn will be successful most of the time: for gaps and conventional CRs, over 90% of the time; for fragments and sluices, over 80% of the time, with a distance of 3 being the next most likely.

3.2.5   Grammatical Analysis

This section proposes basic analyses for the readings that have been identified, and possible approaches for the various forms that allow the relevant readings. This is only a sketch to show how the forms and readings can be handled – detailed analysis is left for chapter 5.

Basic Reading Types

Clausal   Analysis follows G&C: an AVM skeleton for a CONTENT value corresponding to a clausal reading is shown below as AVM (82). It represents a question[49], the propositional content of which is the conversational move made by the source utterance, 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 ... "*).

$$
(82) \quad
\begin{bmatrix}
question \\[2pt]
\text{PARAMS} \quad \left\{\boxed{2}\right\} \text{ or } \{\} \\[6pt]
\text{PROP} \,|\, \text{SOA} \quad
\begin{bmatrix}
illoc\text{-}rel \\
\text{SPKR} \quad \boxed{1} \\
\text{MSG-ARG} \quad \left[\,\ldots\boxed{2}\ldots\,\right]
\end{bmatrix}
\end{bmatrix}
$$

Constituent   Again, analysis follows G&C, as shown in AVM (83). 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.

Again, the abstracted set is shown as either non-empty (containing a parameter corresponding to the queried content, which would be the case for a *wh*-question *"What do you*

---

[49]As introduced in chapter 2, questions are represented as semantic objects comprising a set of parameters PARAMS (empty for a polar question) abstracted over a proposition PROP; they can be thought of as the feature-structure counterparts of λ-abstracts.

mean by 'X'?") or empty (a polar question, as in *"Do you mean X?"*).

$$
(83) \quad
\begin{bmatrix}
\textit{question} \\[4pt]
\text{PARAMS} \quad \left\{\boxed{3}\right\} \text{or } \{\} \\[8pt]
\text{PROP} \mid \text{SOA} \quad
\begin{bmatrix}
\textit{spkr-meaning-rel} \\
\text{SPKR} \quad \boxed{1} \\
\text{SIGN} \quad \boxed{2} \\
\text{CONT} \quad \boxed{3}
\end{bmatrix}
\end{bmatrix}
$$

Lexical   The lexical reading asks about the identity of a word (or phrase) in the source utterance, rather than a part of its semantic content. G&C give no analysis: a proposal is as shown in AVM (84): the content is a question whose propositional content is the proposition that a speaker uttered a string containing a particular word – with this as the parameter being abstracted to form the question. This can therefore be paraphrased as *"Did you say 'X'?"*, or *"What did you say?"*, depending on whether the abstracted set is empty or non-empty:

$$
(84) \quad
\begin{bmatrix}
\textit{question} \\[4pt]
\text{PARAMS} \quad \left\{\boxed{2}\right\} \text{or } \{\} \\[8pt]
\text{PROP} \mid \text{SOA} \quad
\begin{bmatrix}
\textit{utter-rel} \\
\text{SPKR} \quad \boxed{1} \\
\text{SIGN} \quad \left[\ldots\boxed{2}\ldots\right]
\end{bmatrix}
\end{bmatrix}
$$

Corrections   Corrections can be analysed as a lexical reading with an additional layer of intentional structure: instead of $utter(S, X)$, the proposition queried becomes $intend(S, utter(S, X))$, as shown in AVM (85).

$$
(85) \quad
\begin{bmatrix}
\textit{question} \\[4pt]
\text{PARAMS} \quad \left\{\boxed{2}\right\} \text{or } \{\} \\[8pt]
\text{PROP} \mid \text{SOA} \quad
\begin{bmatrix}
\textit{intend-rel} \\
\text{SPKR} \quad \boxed{1} \\
\text{INTENDED} \quad
\begin{bmatrix}
\textit{utter-rel} \\
\text{SPKR} \quad \boxed{1} \\
\text{SIGN} \quad \left[\ldots\boxed{2}\ldots\right]
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

   If corrections can in fact have clausal, constituent or lexical sub-type, the analysis can follow exactly the same lines, embedding the standard reading as shown above.

Literal Reprises and Fragments

As literal sentences appear only to require a clausal reading, their analysis can be as given by G&S and described in section 2.3.4 above: their clause type specifies that their content is a

question which queries the content of a previous utterance, and as that utterance's content is a conversational move, the clausal reading is derived.

The most common readings of reprise fragments (clausal and constituent) are covered by G&C's analysis (see section 2.3.5 above). Although lexical readings are rare, an analysis is possible following the same lines as the constituent reading (treating the fragment as utterance-anaphoric) given an extra contextual coercion operation which corresponds to the lexical reading given above. This operation must produce, given a source utterance, a context in which that utterance is salient and the MAX-QUD is a question about its identity (which word the speaker actually uttered):

(86)

$$
\begin{bmatrix} \text{CONSTITS} & \left\{ \ldots, \boxed{1}, \ldots \right\} \end{bmatrix} \quad \text{(original utterance)}
$$
$$
\Rightarrow
$$
$$
\begin{bmatrix} \text{CONTEXT} & \begin{bmatrix} \text{SAL-UTT} & \boxed{1} \\ \text{MAX-QUD} & ?\boxed{1}.utter\_rel(a, \boxed{1}) \end{bmatrix} \end{bmatrix} \quad \text{(partial reprise context description)}
$$

WH-Versions

Again, G&C give analyses for both of these forms, in which *wh*-substituted reprise sentences receive a clausal reading, and reprise sluices can take either clausal or constituent. It seems that the constituent reading is unnecessary (or at least rare). The clausal readings of the two forms are entirely consistent with one another, being produced by the same contextual coercion operation (parameter focussing) – the difference being that the full sentential version provides its own content which must be unified with the contextually provided MAX-QUD question, while the sluice has its content entirely provided by context. The difficulty of categorising some forms into one class or the other therefore seems unimportant, as suitable analyses are essentially the same.

However, lexical readings are required by the full sentential version, and possibly by both: if the extra coercion operation proposed above for fragments is used, this would directly provide an analysis for sluices. In order to give the full sentential form a lexical reading, an utterance-anaphoric analysis must be available for full sentences as well as just fragments; this is taken up in chapter 5.

Non-Reprise CRs

Non-reprise CRs will be given an analysis by the standard grammatical meaning of the sentence as derived from the lexicon. G&C give an example for a constituent reading (a *"What do you mean . . . "* sentence), and this approach can be extended as required. As shown in AVM (87) below, in the case of the sentence *"What did you say?"*, the verb *say* will be specified in the lexicon as having a semantic content which corresponds to the lexical CR reading. Similarly the verbs *ask*, *tell* will have semantic content containing an illocutionary relation,

giving the clausal reading, and the verb *mean* will contribute the content of the constituent reading.

For polar questions with lexical and constituent readings, the analysis must allow reference to previous (sub-)utterances in order to form a question concerning their meaning or form. This is achieved by using the utterance-anaphoric approach introduced by G&C for the constituent reading (see section 2.3.5) – whereby a word or phrase of type *utt-anaph-ph* can denote a contextually salient utterance. The non-reprise CR *"Did you say 'peanuts'?"* can then be analysed as denoting a question paraphrasable as *"Did you utter the word 'peanuts'?"*, as shown in AVM (88). Clausal equivalents will not require this step (there is no utterance-anaphoricity inherent in a CR such as *"Are you asking me whether **Bo** is leaving"*), but some constituent CRs will (e.g. *"Who do you mean by 'Bo'?"*).

(87)

$$
\begin{bmatrix}
\text{PHON} & \langle \text{what, did, you, say} \rangle \\
\text{SLASH} & \{\} \\
\text{STORE} & \{\} \\
\text{CONT} & \begin{bmatrix} question \\ \text{PARAMS} \quad \{\boxed{1}\} \\ \text{PROP} \quad\quad \boxed{2} \end{bmatrix}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\text{PHON} & \langle \text{what} \rangle \\
\text{LOC} & \boxed{4}\begin{bmatrix} \text{STORE} & \{\boxed{1}\} \end{bmatrix} \\
\text{CONT} & \boxed{1}\begin{bmatrix} param \end{bmatrix}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\text{PHON} & \langle \text{did, you, say} \rangle \\
\text{SLASH} & \{\boxed{4}\} \\
\text{STORE} & \{\boxed{1}\} \\
\text{CONT} & \boxed{2}\begin{bmatrix} proposition \\ \text{SOA} \quad \boxed{3} \end{bmatrix}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\text{PHON} & \langle \text{did} \rangle \\
\text{SLASH} & \{\boxed{4}\} \\
\text{CONT} & \boxed{3}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\text{PHON} & \langle \text{you} \rangle \\
\text{STORE} & \{\} \\
\text{CONT} & x : spkr(x)
\end{bmatrix}
$$

$$
\begin{bmatrix}
\text{PHON} & \langle \text{say} \rangle \\
\text{SLASH} & \{\boxed{4}\} \\
\text{CONT} & \boxed{3}\begin{bmatrix} utter\text{-}rel \\ \text{SPKR} \quad x \\ \text{SIGN} \quad \boxed{1} \end{bmatrix}
\end{bmatrix}
$$

(88)

$$\begin{bmatrix} \text{PHON} & \langle \text{did, you, say, peanuts} \rangle \\ \text{CONT} & \begin{bmatrix} \textit{question} \\ \text{PARAMS} & \{\} \\ \text{PROP} \,|\, \text{SOA} & \boxed{3} \end{bmatrix} \end{bmatrix}$$

$$\begin{bmatrix} \text{PHON} & \langle \text{did} \rangle \\ \text{CONT} & \boxed{3} \end{bmatrix} \qquad \begin{bmatrix} \text{PHON} & \langle \text{you} \rangle \\ \text{CONT} & x : spkr(x) \end{bmatrix} \qquad \begin{bmatrix} \text{PHON} & \langle \text{say, peanuts} \rangle \\ \text{CONT} & \boxed{3} \end{bmatrix}$$

$$\begin{bmatrix} \text{PHON} & \langle \text{say} \rangle \\ \text{CONT} & \boxed{3}\begin{bmatrix} \textit{utter-rel} \\ \text{SPKR} & x \\ \text{SIGN} & \boxed{2} \end{bmatrix} \end{bmatrix} \qquad \begin{bmatrix} \textit{utt-anaph-ph} \\ \text{PHON} & \boxed{1}\langle \text{peanuts} \rangle \\ \text{CONT} & \boxed{2}\begin{bmatrix} \text{PHON} & \boxed{1} \end{bmatrix} \\ \text{CTXT} \,|\, \text{SAL-UTT} & \boxed{2} \end{bmatrix}$$

Conventional CRs

For conventional CR words and phrases, it seems reasonable to propose that their semantic content be specified entirely in the lexicon. Words such as *"what"*, *"eh"*, *"huh"* and phrases such as *"beg pardon"*, *"come again"* can be given lexical entries that give their meaning as precisely the relevant CR question.

(89)

$$\begin{bmatrix} \text{PHON} & \langle \text{eh} \rangle \\ \text{CONT} & \begin{bmatrix} \textit{ask-rel} \\ \text{ADDR} & a \\ \text{MSG-ARG} & \begin{bmatrix} \textit{question} \\ \text{PARAMS} & \{\boxed{1}\} \\ \text{PROP}\,|\,\text{SOA} & \begin{bmatrix} \textit{spkr-meaning-rel} \\ \text{SPKR} & a \\ \text{SIGN} & \boxed{2} \\ \text{CONT} & \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{CTXT}\,|\,\text{SAL-UTT} & \boxed{2} \end{bmatrix}$$
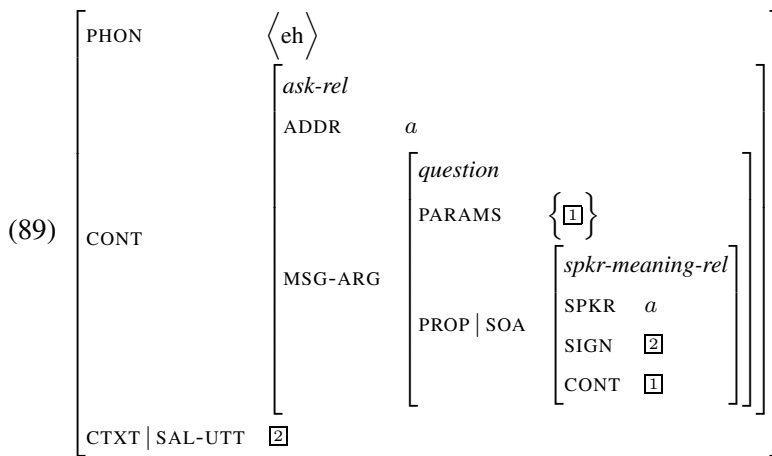
Gaps

The reprise gap form requires a further step, but again seems to be treatable within the same general approach. To capture a lexical reading (the only reading observed), the gap CR fragment itself must be treated as utterance-anaphoric, and a further contextual coercion mechanism defined. In this case, the updated context will be one in which the utterance referred to (and echoed) by the gap CR is salient, and the MAX-QUD is a question about the identity of the word uttered immediately following it:

(90)
$$
\begin{bmatrix} \text{CONSTITS} & \{\ldots, \boxed{1}, \boxed{2}, \ldots\} \end{bmatrix} \quad \text{(original utterance)}
$$
$$
\Rightarrow
$$
$$
\begin{bmatrix} \text{CONTEXT} & \begin{bmatrix} \text{SAL-UTT} & \boxed{1} \\ \text{MAX-QUD} & ?\boxed{2}.utter\_consec(a, \boxed{1}, \boxed{2}) \end{bmatrix} \end{bmatrix} \quad \text{(partial reprise context description)}
$$

Fillers

Fillers could be approached in a similar way, with the same coercion operation operating on the last word uttered in the previous (unfinished) utterance, and thereby producing a contextual question asking about the next word to be uttered:

(91)
$$
\begin{bmatrix} \text{CONSTITS} & \{\ldots, \boxed{1}\} \end{bmatrix} \quad \text{(original utterance)}
$$
$$
\Rightarrow
$$
$$
\begin{bmatrix} \text{CONTEXT} & \begin{bmatrix} \text{SAL-UTT} & \boxed{2} \\ \text{MAX-QUD} & ?\boxed{2}.utter\_consec(a, \boxed{1}, \boxed{2}) \end{bmatrix} \end{bmatrix} \quad \text{(partial reprise context description)}
$$

3.2.6 Conclusions

This section has presented a taxonomy of readings and forms which has been shown to cover nearly 99% of CRs within a corpus of dialogue.

Two of the four readings and four of the eight forms are covered by G&C's HPSG analysis. The remaining readings and forms all seem amenable to similar analyses, and the beginnings of these have been proposed.

Some statistically significant correlations between forms and their possible readings have been shown that will help with disambiguation, although further sources of information will be required for some forms (in particularly the reprise fragment).

The measurements of CSS distance show that an utterance record with length of 4 sentences would be sufficient to allow a dialogue system to process the vast majority of CRs, and that a strategy of considering most recent turns first will be reasonable when disambiguating potential sources.

3.3 Corpus Investigation 2 – Sources

Although the first study provides information about the distribution of different CR forms and readings, it does not provide any information about the specific conditions which prompt particular forms or readings – information which might give a basis for disambiguating ambiguous CRs. It also does not provide information about which word and phrase types are most likely to be clarified (or which CR forms are most likely to be used in each case) – information which will be needed to ensure good coverage by a grammar and/or dialogue system. This section describes a second corpus study, which enriches the first study with further information about the source (word or phrase category, and level of grounding) which might help with both these requirements.

3.3.1 Aims and Procedure

Coverage

G&C's analysis of CRs is only laid out explicitly for proper names, which they take to have a semantic content which fits with this analysis (a parameter with a referential index). It is clear that clarification is possible for other word and phrase types, but it is not clear how the analysis could be extended. One aim of this investigation was therefore to establish which word and phrase types are important, and for which forms and readings: a grammatical analysis can then be produced which gives reasonable coverage. The corpus was therefore re-marked with the word part-of-speech (PoS) category or phrase category of the original source element, for each CR.

Disambiguation

If a system is to correctly interpret user CRs, it must be able to disambiguate between CR forms and readings, and also identify the likely source word or phrase in the original utterance. In addition, if a system is to produce CRs which users can interpret easily and correctly, it must use a sensible CR form for the particular source being clarified and the reading intended.

Disambiguation of Form   Some CRs could be taken as ambiguous between forms: especially in the absence of intonational information, reprise gaps and fragments could be mistaken for one another, as they both take the same surface form (a word or sequence of words echoed from the source utterance). It might be expected that PoS category could be used as a disambiguating factor: for instance, content words (e.g. names, nouns and verbs) might be more likely to have their reference or semantic content queried than function words (e.g. prepositions and determiners), so an echo of a function word might be more likely to be a gap CR than a fragment. Markup of category was hoped to establish whether this kind of effect is really found.

Disambiguation of Reading   Some forms are ambiguous between many readings: the reprise fragment form can convey a constituent, clausal or lexical CR. The factors that influence the production of a lexical CR (or the intpretation of a CR as lexical) are likely to be acoustic or environmental (high noise levels, unclear speech etc.), and as such impossible to examine via this BNC-based corpus. Similarly, disambiguation between clausal and constituent readings may often involve intonation (particular for the reprise fragment form), and again the BNC transcriptions have no intonational data. However, it is likely also to be influenced by semantic and pragmatic factors which may be easier to take into account.

Intuitively, at least two such features might be useful: PoS category and level of grounding. Content words might be more likely to cause constituent queries (concerning their semantic content) than function words, whose meaning could be assumed well known between participants, at least if they share the same language. Similarly, if the source has already occurred and is considered to have been grounded by the participants in a conversation, constituent readings should be less likely. The sentence number (if any) when the source was last mentioned was therefore also marked.

Markup Process

The corpus from section 3.2 was re-examined in order to add this required information. The same set of sentences containing CRs was used, and each one examined in its surrounding context to identify the source element together with any previous mention of that element. The corpus was then re-marked for two attributes: source category and (where applicable) the number of the sentence containing the last previous mention of the source.

A stand-off annotation method (Ide and Priest-Dorman, 1996) was used for this exercise, as opposed to the in-line annotation used previously. This does not fit exactly with the SGML standard used by the BNC, but should make the data easier to re-use with XML-based annotation/analysis systems such as NITE (Bernsen et al., 2002) and MMAX (Müller and Strube, 2003), which are now emerging as more standard.

Reliability of the markup has not been examined. However, the task here is considerably more straightforward than that of section 3.2, so similar or better reliability might be expected.

As before, the markup scheme used for source category evolved during the study and is shown in table 3.12. A full explanation of the different categories (together with subcategories for some classes) is given below in section 3.3.2. Previous mention was marked with the corresponding BNC sentence number, as was done for CSS distance in the previous study.

As when identifying source sentence before, where more than one possible source existed, the most recent was taken.

`unkn`	Source cannot be identified
`uncl`	Source transcribed as `<unclear>`
`utt`	Entire utterance (`<u>` or `<s>` BNC element)
`sent`	Full sentence (possibly sub-utterance e.g. subordinate clause)
`pn`	Proper noun
`pro`	Pronoun
`cn`	Common noun
`np`	Noun phrase (including determiner)
`wp`	*WH*-phrase
`v`	Verb
`vp`	Verb phrase (including argument(s))
`prep`	Preposition
`mod`	Prepositional phrase or other modifier phrase
`conj`	Conjunction
`det`	Determiner (including demonstratives)
`adj`	Adjective
`adv`	Adverb (including polar particles)

Table 3.12: CR source categories

3.3.2 Source Types

Source sentences have already been identified in section 3.2, but exactly identifying which *part* of the source sentence is the element being clarified is more difficult, as is defining what counts as a previous mention of any segment longer than a short fragment (once anaphora, ellipsis etc. are taken into account).

Source Identification

The process of identification of the source was defined as follows. For reprise fragments and literal reprise sentences, the source was taken to be the part of the sentence parallel to the CR (i.e. the echoed part). Example (92) shows a fragment with a definite NP source, example (93)

a literal reprise of a sentential sub-utterance. CRs are shown in **bold**, with sources <u>underlined</u>.

(92)[50]

Gary:	A study into the English language and how it's
Jake:	Oh!
	Oh!
Gary:	spoken in <u>the nineties</u>.
Jake:	**The nineties?**
Gary:	Aha.

(93)[51]

Shelley:	My \<pause\> er, that's it, my sister's
Unknown:	Why are you writing problems?
Shelley:	boyfriend said <u>I'm a common cow and have a got a big nose</u>.
Unknown:	Did she?
	Did he?
Shelley:	Yeah.
Josie:	**You're a common cow with a big nose?**
Shelley:	Yeah.

For sluices and *wh*-substituted sentences, the source was defined as that part corresponding to the *wh*-phrase in the CR. Example (94) shows a sluice with a definite NP source, example (95) a similar *wh*-substituted sentence with an indefinite source:

(94)[52]

Marsha:	who's got <u>their tutor homework</u>?
Carla:	**what tutor homework?**
Marsha:	T P homework

(95)[53]

A.:	you know he, he was, if he, that's, that's possibly <u>one of the nightmare he's having</u> that he doesn't
Arthur:	**That's possibly what?**
A.:	One of the nightmares he's having
Arthur:	What when he's on the drugs, some of these painkillers?

For reprise gaps, two sources were marked: firstly, that part parallel to (echoed by) the repeated word in the CR (as with reprise fragments); secondly, the part immediately following this and apparently being queried. We shall call this second part the *primary* source, as it is the element being queried, and the first part the *secondary* source (see example (96)). In some

---

[50]BNC file KPD, sentences 550–555
[51]BNC file KPG, sentences 485–492
[52]BNC file KP2, sentences 348–350
[53]BNC file KP1, sentences 367–370

cases the primary source may only become obvious after the CR itself – see example (97):

$(96)^{54}$

Laura:	Can I have $\underline{some}_{sec}$ $\underline{toast}_{pri}$ please?
Jan:	**Some?**
Laura:	Toast

$(97)^{55}$

Madge:	erm I think if you're on a committee you have a right $\underline{to}_{sec}$
Anon 5:	**To**
Madge:	$\underline{attend}_{pri}$
Anon 5:	ex- exactly if you're on the committee.

For fillers, again two source categories were marked. In this case, the *secondary* source is taken as the last word from the incomplete utterance; the *primary* source is the new part being queried/suggested by the CR (and which was therefore presumably problematic for the original speaker).

$(98)^{56}$

Jess:	[...] they're always, you know, stand at the back and slag each other off and say oh shit `<unclear>` or stuff but then erm `<pause>` tt afterwards they sort of $\underline{end}_{sec}$ `<pause>`
Catriona:	**$\underline{\textbf{Happy}}_{pri}$?**
Jess:	yeah end happy but, but not, I mean Foxy isn't really `<unclear>` [...]

All examples of non-reprise CRs explicitly queried a particular word or phrase, and this was therefore taken as the source:

$(99)^{57}$

Leon:	Sir, erm, could you please come here , `<laugh>` Luigieans, what?
Unknown:	Comedians.
Leon:	Oh, I thought you said $\underline{Luigieans}$.
Unknown:	**Luigieans, who are they**, `<unclear>`

For conventional CRs, the source (if identifiable and not `unclear`) was taken to be the entire source utterance (the conventional form had already been defined as "indicating complete incomprehension"). In all cases, this seemed suitable.

$(100)^{58}$

Julian:	Now this year might be all right at the end.
Jock:	At the end?
Julian:	$\underline{\text{`<unclear>`} \text{ already.}}$
Jock:	**What?**

In all cases, the surrounding context was read (as in the previous exercise) as it often helped identify what exactly constituted a parallel element; answers to CRs were particularly useful. Except for conventional CRs (whose sources were always marked as whole utterances), the lowest-level definition was taken. For example, if the source consisted of

---

[54] BNC file KD7, sentences 392–394

[55] BNC file KPM, sentences 391–394

[56] BNC file KP6, sentences 294–296

[57] BNC file KPL, sentences 520–523

[58] BNC file KPF, sentences 366–369

a prepositional phrase (PP) which also made up a whole utterance, the category would be marked as PP.

Previous Mention Identification

A word or phrase was judged to be a previous mention of the source word or phrase if it contained the same words in the same order. Other co-referential but not identical expressions are therefore not taken as being previous mentions. This approach seems to be what is desired: in constituent CR cases, we take it that the CR initiator has not understood the reference of the queried phrase, so we cannot assume that she has identified a non-identical but apparently co-referential phrase as a previous mention. It is previous mentions of the actual problematic sequence of words that we are interested in.

In example (101) below, the first occurrence of *"in there"* (shown in [square brackets]) is therefore taken as a previous mention of the CR source (which is the second occurrence of *"in there"*, shown <u>underlined</u> and immediately preceding the CR itself); but the first occurrence of *"in the hospital"* (shown in {curly brackets}) is not taken as a previous mention as it is not word-identical to the source.

(101)[59]

A.:	he panicked and they had a hell of a time with him because he panicked [in there], {in the hospital} and there was a, going berserk and that `<pause>` but you see daddy didn't, oh there's nothing
Unknown:	`<unclear>`
A.:	<u>in there</u>, and he
Arthur:	**In the hospital?**
A.:	Yeah

3.3.3   Results

Table 3.13 shows results taken over all domains, with CR form tabulated against source category; table 3.14 shows the same for CR reading against source category. For gaps and fillers, these figures are for the primary source only.

---

[59]BNC file KP1, sentences 397–400

	uncl	sent	utt	unkn	np	pn	pro	cn	adj	adv	prep	conj	det	mod	v	vp	wp	*Total*
non	3	3	4	1	8	7	7	11	0	0	1	0	0	2	0	1	1	49
lit	2	23	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	28
sub	7	0	0	0	1	2	1	3	1	0	0	0	0	0	0	0	0	15
slu	5	4	0	0	15	5	17	2	1	0	0	0	1	4	0	0	0	54
frg	6	1	0	0	25	21	14	24	3	1	0	0	9	11	0	6	0	121
gap	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	2
fil	0	0	0	0	1	1	0	5	3	0	0	0	0	0	2	2	3	17
wot	37	0	88	5	0	0	0	0	0	0	0	0	0	0	0	0	0	130
oth	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2
Total	60	31	95	6	51	36	39	46	8	2	1	0	10	17	3	9	4	418

Table 3.13: CR form vs. source category

	uncl	sent	utt	unkn	np	pn	pro	cn	adj	adv	prep	conj	det	mod	v	vp	wp	*Total*
cla	12	27	5	1	37	25	33	23	4	1	0	0	8	15	0	6	0	197
con	3	2	20	0	5	9	5	13	0	0	1	0	0	1	0	1	0	60
lex	45	0	68	5	4	2	0	9	4	0	0	0	0	0	3	2	3	145
cor	0	2	0	0	4	0	1	1	0	0	0	0	1	0	0	0	1	10
oth	0	0	2	0	1	0	0	0	0	1	0	0	1	1	0	0	0	6
Total	60	31	95	6	51	36	39	46	8	2	1	0	10	17	3	9	4	418

Table 3.14: CR reading vs. source category

	uncl	sent	utt	np	pn	pro	cn	adj	adv	prep	conj	det	mod	v	vp	wp	*Total*
gap (sec)	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	2
gap (pri)	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	2
fil (sec)	0	0	0	0	0	1	1	1	2	1	2	3	0	5	0	1	17
fil (pri)	0	0	0	1	1	0	5	3	0	0	0	0	0	2	2	3	17
frg	6	1	0	25	21	14	24	3	1	0	0	9	11	0	6	0	121

Table 3.15: CR form vs. primary and secondary source category

	def	indef	time	*Total*
non	5	3	0	8
sub	0	1	0	1
slu	15	0	0	15
frg	11	13	1	25
fil	0	1	0	1
oth	1	0	0	1
Total	32	18	1	51

Table 3.16: CR form vs. NP sub-category

	def	indef	time	*Total*
cla	24	12	1	37
con	3	2	0	5
lex	3	1	0	4
cor	2	2	0	4
oth	0	1	0	1
Total	32	18	1	51

Table 3.17: CR reading vs. NP sub-category

Coverage

The most common cause of CRs appears to be utterances that are unclear and/or incomprehensible as a whole: the `uncl` and `utt` classes make up 37% of CRs.[60] Noun phrases of various kinds then make up the next most common source: proper names, pronouns, common nouns and determiner-noun NPs together make up 41% of CRs. Whole sentences then make up another 7%, with all other classes counting for less then 5% each.

The analysis outlined so far should, in theory, cope with whole sentences and utterances (via literal reprise or conventional CRs), and with unclear elements (via CRs with lexical readings). However, as far as the other categories are concerned, it is so far restricted to PNs (at least for the clausal and constituent readings). This would restrict coverage to 55% of CRs at best.[61] It is clearly important to extend the coverage, with the various types of nominal (NPs, pronouns and CNs) being most important – including these could extend the coverage to 87%. Verbs and VPs seem to be much less important, covering only 3% of our examples. Adjectives, adverbs, prepositions and conjunctions together also make up less than 3%. A

---

[60]The *uncl* class covers those that are marked as unclear by the BNC transcribers, but many other utterances may have been unclear to the addressee, presumably including those lexical CRs whose source is marked as `utt`, a whole utterance.

[61]This figure includes those CRs for which the source could not be identified (marked as `unkn`), just to give the benefit of the doubt.

sensible analysis might therefore be one which covers the various types of NP, plus perhaps determiners and modifier phrases (in which case coverage could theoretically be extended to 94%). Chapter 4 will take up this issue.

Noun-Verb Distinction   Interestingly, this apparent relative importance of noun clarification as opposed to verb clarification does not appear to be merely an effect of overall word frequency. Using the BNC PoS-tagging scheme, the number of noun and verb occurrences in the sub-corpus overall can be counted. These are shown for comparison in table 3.18 below, firstly as raw token counts (the total number of occurrences of a particular PoS tag), and secondly as stemmed type counts (the number of distinct word stems which appear with a particular PoS tag).[62]

In both cases, these overall frequencies are much more similar for the two PoS classes: the noun:verb count ratio is about 1.3:1 for tokens, and about 3.3:1 for types. In contrast, the ratio for CRs is 40:1 (and a $\chi^2_{(1)}$ test confirms that these figures are as significantly different as they appear in both cases).

	pn	pro	cn	v
CRs	36	39	46	3
General (tokens)	5,057	20,435	24,310	38,060
General (types)	1,415	55	3,258	1,413

Table 3.18: Noun vs. verb frequency

The rarity of verb clarification must therefore be an independent effect. Possible explanations might be a syntactic preference for constructing CRs concerning nouns (although this seems unlikely given that all CR forms are included in these figures, and non-reprise forms in particular might be expected to be equally available for different word or phrase types); differences in semantic representation which make verbs more difficult to clarify; differences in the amount of information actually carried by the two classes; or differences in the likelihood that the classes are mutually known in a linguistic community (although one might expect these last two to be reflected in type counts).

Focus   Of those classed as modifier phrases the majority (14 out of 17) were prepositional phrases consisting of preposition and NP. In most cases, the CR seemed to be clarifying the sub-constituent NP more than the PP as a whole. In example (101), repeated here as example (102), the question seems to concern the referent of *"there"* (whether it refers to

---

[62]Stemming was performed using the Porter stemmer (Porter, 1980) with a few minor enhancements allowed by the BNC PoS-tagging scheme – spotting the various forms of the irregular verbs *do*, *be*, *have* and comparative/superlative forms of adjectives.

"the hospital"), rather than, say, a question about the exact location that *"in there"* refers to.

(102)[63]

A.:	he panicked and they had a hell of a time with him because he pan-
	icked <u>in there</u>, <u>in the hospital</u> and there was a, going berserk and that
	`<pause>` but you see daddy didn't, oh there's nothing
Unknown:	`<unclear>`
A.:	<u>in there</u>, and he
Arthur:	**In the hospital?**
A.:	Yeah

If this is really the case, it may be that an analysis for PPs *per se* is not strictly necessary, but rather one which allows sub-constituent NPs to be focussed and clarified. The same may be true for VPs, where reprises seem to be able to focus in on NP arguments in a similar way (in example (103), the question seems to be more whether *"it"* refers to *"maths"*, rather than e.g. whether the maths has been done).

(103)[64]

Carla:	I have done it, oh I haven't, I haven't brung it in
Marsha:	well
Carla:	well he knows I've <u>done it</u> anyway
Marsha:	**done maths?**
Sarah:	homework
Carla:	no I
Sarah:	done your maths?

If so, this could mean that an analysis which includes NPs, nouns and determiners, together with a theory of focussing sub-constituents, could have theoretical coverage as high as 96%. Again, this is discussed more fully in chapter 4.

Content-Function Distinction

Results show that almost all CR sources are content words (or whole phrases or utterances): only 11 (2.6%) were function words, all but 1 of which were determiners (mostly numbers). This is interesting, as it suggests that function words are unlikely to be the source of clarification. This is not merely a symptom of overall frequency of occurrence: in the sub-corpus as a whole, and going by the BNC PoS-tagging scheme, the ratio of content:function word token counts is about 2:1, whereas the ratio of content:function word CR counts is 12:1 (see table 3.19) – in other words, function words in general occur too often to account for the rarity of their CRs. This is even true when examining determiners only (as all function-word CRs except 1 had determiner sources), with the content:function word token counts then being about 5:1. In both cases, the difference from the CR ratio is statistically significant ($\chi^2_{(1)}$ gives a probability $< 0.1\%$, or $< 0.2\%$ for determiners only).

---

[63]BNC file KP1, sentences 397–400
[64]BNC file KP2, sentences 353–359

However, the counts for types (rather than tokens) look rather different, with a content:function word type count ratio of 25:1, or 42:1 when including determiners only – even higher than for CRs. This difference between the type and token distributions shows (unsurprisingly) that function words have a much higher number of tokens per type than content words (about 10 times as many), suggesting that particular function words are more likely to be mutually known and/or carry less information. This may therefore explain the rarity of CRs to some extent, although as the type ratios and CR ratios are still significantly different (confirmed by a $\chi^2_{(1)}$ test) this may not be the whole story.

	Content (pn, pro, cn, adj, adv, v)	Function (conj, det, prep, comp)	Function (det only)
CRs	134	11	10
General (tokens)	109,298	48,752	22,292
General (types)	7,899	319	190

Table 3.19: Content vs. function word frequency

We can therefore take it that function words are unlikely to form the primary source of CRs (although they might form the secondary source of gaps and fillers – see below), and that this might be because their role in sentences is more structural than content- or information-carrying, or perhaps just because their meaning can be assumed mutually known by participants. However, numbers might be worth treating differently from other determiners as they do seem more likely to be sources. It should also be noted that this general effect may not hold for specific domains: for example, in task-oriented dialogue where, say, the difference between placing one construction part *over* or *under* another may be crucial, prepositions might well be more likely to be clarified than in the general conversation reflected in the corpus being used here.

Form-Source Correlation

For whole utterances or sentences, choice of CR form seems straightforward: utterances appear to be most often clarified by conventional CRs (93% of cases), and sentences by literal reprises (74% of cases). Conversely, identifying the source of these CR forms seems straightforward: all literal reprises except one queried a whole sentence, utterance or unclear segment, as did 96% of conventional CRs.

For NPs and CNs (which we have seen are the other important classes), more forms are available: table 3.13 shows that the fragment, sluice and non-reprise forms are all common, with *wh*-substituted sentences also being possible. For the NP and pronoun categories, there is no strong preference for any particular form, but both PNs and (even more strongly) CNs show preferences.

Fragments vs. Sluices   About 58% of PN sources cause a reprise fragment CR: with PNs, fragments are about 4 times more likely than reprise sluices, and 3 times more likely than non-reprise CRs. Over 50% of CN sources cause a fragment, with fragments twice as likely as non-reprise CRs, and 8 times as likely as sluices. This is statistically significant in both cases: comparing the numbers of fragments vs. sluices against those for NPs and pronouns combined gives a probability of independence of just over 2% for PNs and $< 0.1\%$ for CNs, according to a $\chi^2_{(1)}$ test. This means that this can be used when disambiguating user CRs: sluices are unlikely to be querying a PN or CN source (only 13% of them do) if possible NP or pronoun sources are available (59% of sluices query sources of these classes).

Definites vs. Indefinites   Closer examination of NPs, however, does show an interesting pattern. Tables 3.16 and 3.17 show the figures for NPs, broken down further into definites, indefinites and one other which did not easily fit either classification (*"five past two"*). While there are no clear differences in reading between definites and indefinites, the figures for form show that reprise sluices do not appear with indefinites ($\chi^2_{(1)}$ gives a probability of independence $< 0.1\%$ when comparing fragments vs. sluices). This suggests a difference in the clarificational potential of definites and indefinites, and we will take this up again in chapter 4.

Fillers and Gaps   Table 3.15 shows the figures for primary and secondary sources, for fillers and gaps, together with the standard figures for fragments. For both fillers and gaps, we might expect to see that function words can be secondary sources (words that come immediately before the problematic element) but not primary sources (the actual element being clarified). If so, this would help us disambiguate user input: elliptical utterances consisting of function words could be taken to be gaps rather than reprise fragments.

For gaps themselves, there is not enough data to draw any firm conclusions, but for fillers perhaps we can. 30% of secondary sources are function words, whereas no primary sources are. This is statistically significant: comparing the function-word/non-function-word distribution between secondary and primary filler sources shows a significant difference, with $\chi^2_{(1)}$ giving a probability of independence of 1.5%. In addition, comparing this distribution between secondary filler sources and fragments shows a significant difference ($p = 0.5\%$), whereas comparing it between primary filler sources and fragments shows no difference ($p = 24\%$). It therefore seems that secondary sources behave differently from standard CRs (i.e. fragments), and from primary sources, in that they can be function words. This gives some confidence in our prediction, but more data is required, especially for gaps.

Reading-Source Correlation

Only 1 of the 11 function words seemed to have a constituent reading, with most having a clausal reading, and this does fit with the original expectation that function words would be unlikely to get constituent readings. However, the small number of cases means that this effect is not statistically significant – the probability that the content/function distinction and the clausal/constituent distinction are independent is over 50% according to a $\chi^2_{(1)}$ test.

Markup of last mention of the original source fragment has also not given results with any level of confidence. It is reprise fragment CRs for which we are most interested in being able to disambiguate reading, and for these all constituent readings do occur on the first mention of the source fragment (as expected) – but there are too few of these examples (7) to draw any firm conclusions. It is also impossible to know whether first mention in the transcription is really the first mention between the participants: we do not know what happened before the tape was turned on, what their shared history is, or what is said during the frequent portions marked as `<unclear>`.

So the current corpus cannot provide enough information for this purpose. In order to examine these effects properly, a new experimental technique was designed which allows dialogues to be manipulated directly, with CRs with the desired properties automatically introduced into the conversation. Section 3.5 describes this technique and the experiment performed.

Clarifying CRs

The information about source that this exercise has provided allows a further question to be answered: is it possible to clarify a CR itself, and if so, how often does this happen? In principle, it would seem that it should be possible to clarify a CR in the same way as any other utterance, if the hearer has not understood the reference of some part of it (e.g. the reference of a NP in a clausal reprise fragment, or even the utterance which an utterance-anaphoric constituent or lexical question is intended to ask about). However, if such CRs-of-CRs do *not* occur, this might allow a simpler grammar (as CRs themselves might not need the contextual abstract representation proposed for most utterances) and perhaps a simpler information state representation (as only one utterance could be under clarification at any time). Conversely, if CRs-of-CRs are common and can occur in long stacked sequences, this may also influence our IS representation and the assumptions underlying ellipsis resolution and grounding protocols.

This investigation produced 8 examples of CRs whose source was part of another CR – see example (104). In 5 cases, the source was a non-reprise CR, and in the other 4, the source was a correction. In most cases, the source word/phrase itself within the first CR was a NP of some kind (definite, indefinite or pronoun), but in one case was a (number) determiner and in one case a whole utterance. No cases of more than one level were seen (e.g. CRs-of-CRs-of-

CRs).

	Andy:	what, whe- where does the priest go, or priests go, he's quite young, and he goes walking, he goes, have you really got <u>S and M</u> then?
(104)[65]	Monica:	**What's <u>that</u>?**
	Andy:	**S and M?**
		Sadomasochism.

While it seems that there may be a pattern in that only non-reprise and correction CRs have been seen to provide sources, the numbers are too small to draw definite conclusions from this. We can conclude, though, that CRs-of-CRs exist, and that a grammar and dialogue system should take this possibility into account. They are clearly not common from an overall viewpoint: 8 examples makes up less than 0.1% of the sentences in the corpus. However, 8 of 418 cases is about 2% of the CRs observed, which is at least of the same order of magnitude of the 3% of sentences in general that are clarified, and this suggests that CRs are roughly as likely to be clarified as sentences in general are. Perhaps, then, the grammatical representation of CRs should allow for clarification in just the same way as that of standard sentences. As for stacks of multiple-level CRs, as none have been seen it seems safe to assume that they are not especially common, although we have no evidence as to whether they can exist or not.

3.3.4   Conclusions

This section has presented an investigation into the nature of the sources of CRs. We have seen that an analysis which treats CNs and various types of NPs as possible CR sources is vital to give good coverage, but that other categories such as verbs and function words are not important; we have seen that CRs themselves can function as sources of clarification; and we have seen some significant correlations between source type and CR form:

- Most CRs ask about whole sentences and utterances, or about NPs of various kinds. Verbs and function words do not often form CR sources (although this might change in specific domains).

- Conventional CRs usually ask about whole utterances.

- Literal reprises usually ask about whole sentences.

- *WH*-substituted reprises, reprise fragments, reprise sluices and non-reprise CRs all usually ask about NPs, with reprise sluices likely to ask about definite NPs or pronouns in particular.

We have also seen suggestions that function words, while unlikely to act as primary CR sources, can act as secondary sources for gaps and fillers; that when function words do act as

---

[65]BNC file KPR, sentences 464–467

primary sources the CR may be unlikely to have a constituent reading; and that constituent readings may be more likely when the source is a first mention – but for all of these we require more data to draw firm conclusions.

3.4   Corpus Investigation 3 – Responses

In order to treat CRs fully within a dialogue system, we must ensure that the system is capable of correctly processing the types of answers that a user may give. It will also be important to ensure that the system answers any user CRs in a suitable manner and at a suitable time. This section describes a third corpus study, which investigated the nature of responses to CRs (whether they are answered, how they are answered, and when the answers come), and discusses the implications for a CR-capable dialogue system.[66]

3.4.1   Aims and Procedure

Answering User CRs

If a system is to be able to deal with user-generated CRs, it must respond in an appropriate manner. The type of answer must depend on the question being asked by the CR: it seems clear that a polar yes/no answer might be appropriate for a yes/no question (such as that asked by a clausal reprise fragment), but not for a *wh*-question (such as that asked by a constituent equivalent). The suitability of other features of the answer (e.g. elliptical or non-elliptical) may also depend on both form and reading. In order to get information to help establish the requirements for a system which must generate answers to CRs, the corpus was re-marked for the type of answer seen for each CR.

Answering System CRs

The same applies when processing user answers to system CRs, but in this case having information about expected answer types is even more important, as there is an extra question: whether a user response is really an answer to a system CR or is in fact some other unrelated move. If we know what type of answer a particular CR is likely to receive, and indeed whether it is likely to receive an answer at all, this will help disambiguate a user move (particularly an elliptical fragment) between being an answer to the CR or, say, an answer to a different question, or even a new user CR. In addition, the likely distance between CRs and answers will help with this – if all answers to CRs come immediately, this saves having to consider later user moves as possible answers to a distant CR. The location of answers (the BNC sentence number) was therefore also marked, allowing distance between CRs and answers to be calculated.

---

[66]Much of the work in this section has appeared in (Purver et al., 2003b).

Crossed/Nested Answers

Determining when CRs are answered also provides information on whether CRs are addressed in a strictly stack-like fashion, or whether CR-answer pairs can cross each other. The utterance-processing protocol proposed by G&C assumes that utterances under clarification are kept in a stack structure, with only the most recent being available for subsequent grounding or accommodation. Similarly, both GoDiS and IBiS assume a stack of questions under discussion, with only the most recently raised question available to be answered and resolved.

If this is the case, dialogues such as the imaginary example (105) should not be possible, where question-answer pairs are crossed (A(5) is taken to answer B(2), and A(7) to answer B(4). Here CRs are again shown **bold**, with answers <u>underlined</u>). Dialogues such as example (106) (again imaginary), where multiple levels of clarification exist but are nested within each other so that the most recent unanswered one is always used to interpret the answer, should be fine.

(105)

A(1):	Did Bo leave yesterday?
B(2):	**Yesterday?**
A(3):	I was just wondering if you knew.
B(4):	**Who is Bo?**
A(5):	<u>Yes, yesterday lunchtime</u>.
B(6):	Oh, right.
A(7):	<u>Bo Smith, the linguist</u>.
B(8):	Oh yes, I know her. Yes, she did.

(106)

A(1):	Did Bo leave yesterday?
B(2):	**Yesterday?**
A(3):	I was just wondering if you knew.
B(4):	**Who is Bo?**
A(5):	<u>Bo Smith, the linguist</u>.
B(6):	Oh yes, I know her.
A(7):	<u>Yes, yesterday lunchtime</u>.
B(8):	Oh, right. Yes, she did.

Marking the sentence numbers of answers allows any crossed or nested pairs to be spotted directly.

Markup Process

Again, the corpus from section 3.2 with the same set of CRs was used, and each one examined in its surrounding context to see if a response could be identified. The corpus was then re-marked for two attributes: response type and (where applicable) the number of the sentence containing that response.

11 CRs in the previous corpus were self-clarifications and therefore not considered during

this study (as they would not be responded to in the same way as others, and were not likely to be relevant to dialogue system design), and one was also excluded as it was part of a telephone conversation of which only one side was transcribed. The total number of CRs in the study was therefore 406. As in section 3.3, stand-off annotation was used, and reliability has not been examined; again, as the task here is more straightforward than that of section 3.2 it could be expected to be similar or better.

As before, the markup scheme used for response type evolved during the study and is shown in table 3.20. A full explanation of the different types is given below in section 3.4.2.

uncl	Possible answer but transcribed as `<unclear>`
cont	CR initiator continues immediately
none	No answer
qury	CR explicitly queried
frg	Answered with parallel fragment
sent	Answered with full sentence
yn	Answered with polar particle

Table 3.20: CR response types

Answer location was marked in terms of the sentence numbering scheme in the BNC, just as was done in the previous sections for source and previous mention – CR-answer separation distance (hereafter CAS distance) was then calculated from this, both in sentences and speaker turns.[67]

3.4.2 Answer Types

The first three categories correspond to different types of apparently unanswered CRs: firstly, those that may have been answered, but where the sentence possibly containing an answer was transcribed in the BNC as `<unclear>`, as shown in example (107); secondly, those that appear to have remained unanswered because the CR initiator continued their turn without pause, as shown in example (108); and finally those that are not explicitly answered at all (or at least where we have no indication of an answer – eye contact, head movement etc. are not recorded in the BNC but could function as answers), as in example (109).

(107)[68]
Rob:	`<unclear>` try to get that lad who was up at `[name]` with a machine.
Mick:	**Who Malcolm?**
Unknown:	<u>`<unclear>`</u>
Mick:	`<unclear>` get a job at er Oxford.

---

[67] As with CSS distance in section 3.2, although a measure incorporating discourse structure might give even more information, this simple sentence/turn measure seems sufficient for the current purposes.

[68] BNC file KP7, sentences 362–364

(108)[69]
| Marsha: | . . . yeah, yeah, you see the `<unclear>` on her face |
| Carla: | **who Sukey?**, she goes to me, he I doubt, doubt if he'd fancy me, but she's just like, she's all going I doubt if he fancy me and do you think he does? |

(109)[70]
Anon 1:	Guess what, they made one for Super Nintendo.
Selassie:	**Tazmania?**
Anon 1:	It's so cack `[sic]` When you're doing the running, you're running from behind, you can only see the behind of Taz `<pause dur=3>` you see the behind of Taz

A further type was used to classify cases where the CR is explicitly responded to, but by asking another question concerning the CR rather than providing an answer:

(110)[71]
Daniel:	Why don't you stop mumbling and
Marc:	Speak proper like?
Daniel:	speak proper?
Unknown:	**Who?**
Daniel:	Who do you think?
Unknown:	You.

The remaining types represent cases where the CR is explicitly answered, either by a polar particle (*yes*, *no*, *possibly* etc.) as shown in example (111), by a full sentence as shown in example (112), or by an elliptical fragment as shown in example (113).

(111)[72]
cd:	. . . erm two milli-kelvins from absolute zero.
sb:	That's about two thousandth
cd:	Two thousandth
sb:	erm degrees. Yes.
cd:	**Two thousandth of a, of a degree**
sb:	Yes.

(112)[73]
Peter:	Why are you in?
Danny:	**What?**
Peter:	Why are you in?
Danny:	Drama.

---

[69]BNC file KP1, sentences 256–257
[70]BNC file KNV, sentences 546–548
[71]BNC file KNY, sentences 315–320
[72]BNC file KRH, sentences 1019–1025
[73]BNC file KPT, sentences 469–472

(113)[74]
Unkown:	Do you know Adam?
Skonev:	**Adam Adam `[name]`?**
Unkown:	<u>Adam `[name]`</u>

In some cases, the initial response was followed by further information, as shown in examples (114) and (115). In these cases both the initial response type and the type of the subsequent material were recorded (example (114) was classified as `yn+sent`, example (115) as `frg+sent`). The initial type was taken as the main (*primary*) type and used to classify the response for the main results below; the subsequent *secondary* responses are also examined briefly.

(114)[75]
Unknown:	Will you meet me in the drama studio?
Caroline:	**Drama studio?**
Unknown:	<u>Yes</u>$_{pri}$ <u>I've got an an audition</u>$_{sec}$.

(115)[76]
Craig:	He ain't even seen a fight though has he?
Jill:	**Who?**
Craig:	<u>That bloke</u>$_{pri}$, <u>he's got two massive `<unclear>`</u>$_{sec}$.

3.4.3   Results

Response Type

Results for response type are shown in table 3.21 as raw numbers, and also summarised in table 3.22 as percentages for each CR type, with the `none`, `cont`, `uncl` and `qury` classes conflated as one "unanswered" (`unans`) class, and only the most common 4 CR forms shown.

The most striking result is perhaps the high overall number of CRs that do not receive an explicit answer: 39% of all CRs do not appear to be answered overall. This figure is reduced to 17% when taking account of those marked `uncl` (possible answers transcribed as `<unclear>`) and `cont` (the CR initiator continues without waiting), but this is still a significant number. Interestingly, the most common forms (conventional and fragment) appear to be answered least – around 45% go unanswered for both. The form which appears to be most likely to be answered overall is the non-reprise form.

This may partly be caused by the fact that most of the BNC dialogues are transcribed from face-to-face conversations: answers can be given by gestures, eye contact etc. which will not be recorded in the transcription. It may also be due to the general nature of most of the dialogues: in task-oriented dialogues where successful understanding of each utterance is more important, one might expect more CR answers to be found. It seems likely though

---

[74]BNC file KR1, sentences 310–312
[75]BNC file KP3, sentences 936–938
[76]BNC file KP9, sentences 465–467

	none	cont	uncl	qury	frg	sent	yn	*Total*
non	4	2	1	0	14	12	12	45
lit	5	2	1	0	1	2	15	26
sub	4	0	3	0	4	4	0	15
slu	7	6	5	1	27	8	0	54
frg	22	22	6	0	25	4	37	116
gap	1	0	0	0	1	0	0	2
fil	4	0	0	0	7	1	5	17
wot	22	13	25	0	11	57	1	129
oth	0	0	0	1	0	1	0	2
Total	69	45	41	2	90	89	70	406

Table 3.21: Response type vs. CR form

	unans	frg	sent	yn	*Total*
non	15.6	31.1	26.7	26.7	(100)
slu	35.2	50.0	14.8	0.0	(100)
frg	43.1	21.6	3.4	31.9	(100)
wot	46.5	8.5	44.2	0.8	(100)

Table 3.22: Response type as percentages for each CR form

that in some cases, an explicit answer is not found either because an answer is not expected
(in the case of CRs produced purely for grounding or acknowledgement purposes) or because
continuation on the same subject functions as an implicit (affirmative) answer. This may be
the case in example (109) above, for example. This seems to be borne out by the fact that CRs
with clausal and lexical readings go unanswered more often than constituent questions (see
below), which seem much less likely to be used for grounding/acknowledgement purposes.

This indicates that a dialogue system which produces CRs must be prepared for them not
to be answered: when producing grounding or check questions, non-contradictory continua-
tion should be taken as an implicit answer. When producing content questions which require
answers, a non-reprise form may help increase the likelihood of a response.

	none	cont	uncl	qury	frg	sent	yn	*Total*
cla	32	31	11	2	47	15	58	196
con	9	3	0	0	20	20	6	58
lex	22	11	30	0	22	53	6	144
cor	1	0	0	0	1	0	0	2
oth	5	0	0	0	0	1	0	6
Total	69	45	41	2	90	89	70	406

Table 3.23: Response type vs. CR reading

	unans	frg	sent	yn	*Total*
cla	38.8	24.0	7.7	29.6	(100)
con	20.7	34.5	34.5	10.3	(100)
lex	43.8	15.3	36.8	4.2	(100)

Table 3.24: Response type as percentages for each CR reading

Form/Response Correlation

Some CR forms appear to have high correlations with particular response types, giving a good idea of how sluices, conventional CRs and literal reprises should be answered. As might be expected, sluices (which are *wh*-questions) are generally answered with fragments, and never with a polar yes/no answer; less obviously, using full sentences to answer sluices also seems rare (fragments are used 77% of the times they are answered). Yes/no answers also seem to be unsuitable for the conventional CR form, which is generally answered with a full sentence (83% of the times they are answered). Literal reprise sentences are generally answered with polar yes/no answers (83% of the time) rather than with a full sentence (e.g. repeating the echoed sentence a third time).

Reprise fragments, however, while not often answered with full sentences, can be responded to either by fragments or yes/no answers. This may still be useful in a dialogue system: as long as ellipsis reconstruction can process a fragment or yes/no answer to a question, this is likely to result in a direct answer. In contrast, indirect answers (which may require inference, domain knowledge etc. to process and determine their relevance) would be more likely to be given in the form of full sentences. If we want to avoid indirectness and its associated difficulties, using a CR form which is likely to result in a fragment or polar answer may be desirable – so using reprise fragments, sluices or literal reprises where possible is likely to be more useful than using a conventional CR.

Non-reprise CRs seem to be able to be answered using any form – this might be expected given the range of possible syntactic forms and questions encompassed by this class. For the gap, filler and *wh*-substituted forms, no significant conclusions can be drawn.

Reading/Response Correlation

From tables 3.23 and 3.24 (again, percentages given for each CR reading, with "unanswered" response types conflated and only the most common 3 readings shown) we can see that there is a correlation between reading and response type, but that this correlation is also not as simple as a direct reading-answer correspondence. Clausal CRs are unlikely to be answered with full sentences, but can get either fragment or yes/no responses. Constituent CRs are less likely to get yes/no responses but could get either other type. Interestingly, constituent CRs seem to be roughly twice as likely to get a response as clausal or lexical CRs (even though there are fewer

examples of constituent CRs than the others, this difference is statistically significant, with a $\chi^2_{(1)}$ test showing $< 0.5\%$ probability of independence); this might be a result of clausal and lexical CRs being more likely to be used for grounding/acknowledgement purposes as suggested above.

Reprise Fragments   The results in table 3.21 showed that reprise fragments showed no general preference between fragment or yes/no answers. However, examination of the results for this form when taking the reading into account does show a preference: as shown in table 3.25 (raw figures) and table 3.26 (percentages for only those examples with responses and for the two most common readings), the clausal and constituent versions tend to be answered in different ways (as might be expected given G&C's analysis which treats the content of the clausal version as a polar question and the content of the constituent version as a *wh*-question). Clausal versions are more usually answered with a polar answer (61% of the time), whereas constituent versions are much more likely to get fragments as answers (80% of the time). While the amount of data is small, the difference in distributions is significant ($p < 5\%$ according to a $\chi^2_{(1)}$ test).

	none	cont	uncl	qury	frg	sent	yn	*Total*
cla	16	22	6	0	19	4	36	103
con	2	0	0	0	4	0	1	7
lex	0	0	0	0	1	0	0	1
cor	0	0	0	0	1	0	0	1
oth	4	0	0	0	0	0	0	4
Total	22	22	6	0	25	4	37	116

Table 3.25: Response type vs. CR form (fragments only)

	frg	sent	yn	*Total*
cla	32.2	6.8	61.0	(100)
con	80.0	0	20.0	(100)

Table 3.26: Response type as percentages for each CR reading (fragments only)

Answer Distance

Results for CR-answer distance are shown in tables 3.27 and 3.28 (including the `uncl` and `qury` classes). It is clear that the vast majority (94–95%) of CRs that are answered are answered in the immediately following sentence or turn, and that none are left longer than 3 sentences (or turns). The two cases where distance in turns is zero are both cases where a speaker answers their own question explicitly.

Distance	0	1	2	3	>3	Total
Sentences	0	275	15	2	0	292
Speaker Turns	2	278	10	2	0	292

Table 3.27: Number of CRs vs. CAS distance

The same pattern appears to hold across the different response types: 89% of polar answers have a CAS distance of 1 sentence (91% with 1 speaker turn); 94% of fragments (96%); and 97% of full sentences (98%).

Response Type		0	1	2	3	>3	Total
`yn`	*(sentences)*	0	62	8	0	0	70
`yn`	*(turns)*	1	64	5	0	0	70
`frg`	*(sentences)*	0	85	4	1	0	90
`frg`	*(turns)*	1	86	2	1	0	90
`sent`	*(sentences)*	0	86	2	1	0	89
`sent`	*(turns)*	0	87	1	1	0	89

Table 3.28: Response type vs. CAS distance

This contrasts strongly with figures for non-clarificational questions – table 3.29 shows figures taken from (Fernández et al., forthcoming), a study of elliptical utterances in general on a similar dialogue sub-corpus of the BNC. We can see that elliptical fragment answers to questions in general are much less immediate: only 55% have a distance of 1 sentence, with some distances above 10 sentences. The difference between this distribution and that for `frg` CR answers is significant, with the probability of independence from a $\chi^2_{(3)}$ test $p < 0.1\%$. Polar answers, on the other hand, do seem much more similar, with 95% having a distance of 1 sentence (there is no statistically significant difference between this distribution and that for the `yn` CRs, with $\chi^2_{(3)}$ giving a probability of 10%). In fact, the distribution for CRs overall (and even for `frg` CR answers taken on their own) is very similar to that for normal polar answers, with $\chi^2_{(3)}$ giving a probability of independence of nearly 30% (above 50% for `frg`s). Answers to CRs therefore seem to behave more like normal polar answers, even if they are fragments.

Distance	0	1	2	3	>3	Total
Short Answers (sentences)	0	104	21	17	46	188
Affirmative Polar Answers (sentences)	0	104	4	0	1	109

Table 3.29: Answer distance for elliptical answers in general

We can therefore conclude that (a) answering user CRs must be done immediately, and that any dialogue management scheme must take this into account, and (b) we should expect

answers to any system CRs to come immediately – interpretation routines (and this may be especially useful for any ellipsis resolution routines) can assume that later turns are relevant to something other than the CR (if there are other candidates).

Nesting/Crossing

No examples of crossed CR-answer pairs were found: two examples do appear to be crossed based on BNC sentence number alone, but closer examination of the BNC markup details shows that in both cases, two consecutively numbered sentences are actually spoken simultaneously, removing the apparent crossing effect. This investigation therefore provides no counter-evidence to a stack-based dialogue processing and grounding protocol (although it doesn't of course rule out the possibility that such counter-evidence might be found given more data).

Only one example of a possible nested pair was observed, shown above as example (104) and repeated here as example (116). Here, the nested CR *"S and M?"* is not overtly answered either within or outside the other pair, but Andy's continuation suggests that the nested CR has been answered affirmatively by non-verbal means, or is just assumed to be answered:

(116)[77]

Andy:	what, whe- where does the priest go, or priests go, he's quite young, and he goes walking, he goes, have you really got S and M then?
Monica:	**What's that?**
Andy:	**S and M?**
	Sadomasochism.

The scarcity of nested pairs suggests that a stack-based dialogue processing protocol is likely to involve only small numbers of concurrently pending utterances (and therefore that the stack is unlikely to grow large enough to cause any problems). The fact that they seem to be possible confirms that a stack is indeed required, rather than a single variable.

Secondary Responses

Examination of secondary responses showed that they are not as common as might be expected. For sentential responses (table 3.32) they hardly happen, with only one example of a sentence followed by a polar particle found. Fragment responses (table 3.30) are unlikely to be followed by further material — 87% are not — but both sentences and yes/no answers are possible. While no yes/no secondary answers are seen with constituent questions, as might be expected, there is no significant difference between the distributions for the various readings. Yes/no answers (table 3.31) are the most likely type to be followed by secondary answers, but still 67% are not.

---

[77]BNC file KPR, sentences 464–467

	frg	frg+sent	frg+yn	*Total*
cla	39	6	2	47
con	17	3	0	20
lex	20	1	1	22
Total	77	10	3	89

Table 3.30: Secondary response type vs. CR reading (fragments)

	yn	yn+sent	yn+frg	yn+uncl	*Total*
cla	43	8	6	1	58
con	3	3	0	0	6
lex	1	5	0	0	6
Total	47	16	6	1	70

Table 3.31: Secondary response type vs. CR reading (yes/no answers)

	sent	sent+yn	*Total*
cla	14	1	15
con	20	0	20
lex	53	0	53
Total	88	1	88

Table 3.32: Secondary response type vs. CR reading (sentences)

This seems to contrast with Hockey et al. (1997)'s findings based on the Map Task corpus (Anderson et al., 1991) that only 40% of answers to check questions which included a yes/no particle were bare; check questions are questions "requesting confirmation of information that the checker has some reason to believe but is not entirely sure about", so should correspond more or less to the clausal CR reading here, at least for reprise fragments and literal reprise sentences. If these forms are examined individually, the results are very similar to those above (e.g. for reprise fragment CRs with a clausal reading, 70% of yes/no answers are bare). However, it does agree well with their findings for other general yes/no questions – they found that 64% of answers which included a yes/no particle were bare.

This suggests that for a dialogue system, giving primary answers only will be suitable in most cases. It may be worth supplementing some yes/no answers, however (e.g. following Wahlster et al., 1983), as these get further material in about a third of cases. Further corpus investigation might be helpful here – Hockey et al. (1997) saw that the likelihood of bare answers depended on whether the answer was affirmative or negative, and on answer expectations.

3.4.4 Conclusions

This section has presented an investigation into how and when CRs in a dialogue corpus are answered. Some strong correlations between CR form and expected answer type are clear, and can be used in a system both to answer user CRs in a natural manner and to help process user responses to system CRs. In particular:

- Conventional CRs should be answered with full sentences.

- Reprise sluices, and reprise fragments with constituent readings, should be answered with fragments.

- Literal reprise sentences, and reprise fragments with clausal readings, should be given yes/no answers.

- In most cases, simple answers are fine, but some consideration should be given to supplementing yes/no answers with further information.

We have also seen that responses to CRs, when they come, come immediately, but that many CRs do not appear to receive responses at all (at least in the corpus examined here).

3.5 Experiments

This section describes an experimental setup that was designed, together with some software, to allow experiments into specific dialogue phenomena to be performed by introducing them into a real dialogue, and presents the results of an experiment using this setup to extend the corpus work on CR responses and disambiguation.[78]

3.5.1 Aims and Experimental Design

This experiment concentrated on the reprise fragment and gap forms. The results of the previous sections show that the reprise fragment form is the one we are most interested in being able to disambiguate: it is not only common (approximately 30% of CRs in the corpus) but can appear with all the main readings (although biased towards a clausal reading – 86% of occurrences). It can also be easily mistaken for the reprise gap form, at least when no intonational information is available (say, to a dialogue system that is text-based or lacks accurate pitch contour detection). Section 3.3 gave some indications that both PoS category (particularly the content/function distinction) and level of grounding (previous mention) of the original source are likely to affect the reading and form that people attribute to a fragment, but not enough data was available from the corpus to allow significant conclusions.

---

[78]Much of the work in this section has appeared as (Purver et al., 2003b; Healey et al., 2003).

By their echoic nature, these forms specify their sources (primary or secondary) quite precisely, and therefore allow the effect of features of the source element to be examined. The aim of the experiment was therefore to artificially introduce a number of *echoed fragments* (EFs) — fragments repeated from the previous turn, which could be interpreted as reprise fragments or gaps — into real dialogues, and observe how the responses of the participants (and the forms and readings that these responses implied) varied with PoS and previous mention. Specifically, the following hypotheses were to be tested:

- EFs of function words will be likely to be interpreted as gaps. If interpreted as fragments, they will be given clausal readings. First/second mention should make little difference.

- EFs of content words will be interpretable as fragments with clausal or constituent readings, although more constituent readings might be expected on first mention than on second mention.

Experimental Setup

The basic setup involved pairs of subjects, seated in different rooms, communicating using a text-based chat tool (see below). This tool automatically introduced artificially generated EFs to the dialogues: each one was visible only to one participant (the one whose original turn it was querying), and looked as though it came from the other participant, as shown in figures 3.11 and 3.12.

The use of text-based dialogue and separate rooms was to rule out the possibility of communication by gestures and other non-verbal means, and also to rule out the lexical reading (the dialogue history was visible at all times, so the question of what words had been used should not arise).

The subjects were not told that artificial turns could be generated until afterwards. First they were asked about the naturalness of the dialogue: none had noticed any unnatural features of the dialogue (although a few commented that their partners had asked some stupid questions!). The setup was then fully disclosed and subjects were given the choice to have their data deleted. None elected to do so.

Chattool Software

The chattool software used for this experiment was created especially for the purpose. In overview, it allows pairs of subjects to communicate using a synchronous text chattool (see figure 3.13 for an example). However, instead of passing each completed turn directly to the appropriate chat clients, each turn is routed via a server, which can modify turns in predefined ways. Words or spelling can be changed according to specified rules to control what each participant sees and allow miscommunication to be set up. The display of turns can be

Subject A's View		Subject B's View

B: `I was such a cute`
   `baby, I still am cute`
	⟵	B: `I was such a cute`
		   `baby, I still am cute`

A: `no ur not!!`	⟶	A: `no ur not!!`

B: `Obviously the`
   `relatives were coming`
   `around like they do`
   `to see me`
	⟵	B: `Obviously the`
		   `relatives were coming`
		   `around like they do`
		   `to see me`

	Probe →	A: `relatives?`
	Block ←	B: `Yeah just unts and` `uncles`
	Ack →	A: `ah`
A: `yeah`	⟶	A: `yeah`

B: `One of my uncles who`
   `is experienced from`
   `having five kids`
   `said to my dadthat`
   `i looked pale`
	⟵	B: `One of my uncles who`
		   `is experienced from`
		   `having five kids`
		   `said to my dadthat`
		   `i looked pale`

Figure 3.11: Story Telling Task Excerpt, Noun Clarification, Subjects 1 & 2

Subject A's View		Subject B's View
B: `go on chuck her out`	⟵	B: `go on chuck her out`
A: `so we agree`	⟶	A: `so we agree`
B: `agree?`	← Probe	
A: `yeah to chuck out` `Susie derkins`	→ Block	
B: `uh huh`	← Ack	
A: `yes`	⟶	A: `yes`

Figure 3.12: Balloon Task Excerpt, Verb Clarification, Subjects 3 & 4

allowed or prevented for each participant, and (importantly for this experiment) new artificial turns can be generated automatically according to features of the dialogue.

User Interface

The user interface was written in Java by another project member (James King at Queen Mary, University of London). It is similar to instant messaging applications such as ICQ or MSN Messenger: its window is split into two panes, a lower pane used to enter text, and an upper pane in which the conversation is displayed as it emerges (see figure 3.13). A status display between the two panes shows whether the other participant is active (typing) at any time: this can be artificially controlled during the generation of artificial turns to make it appear as if
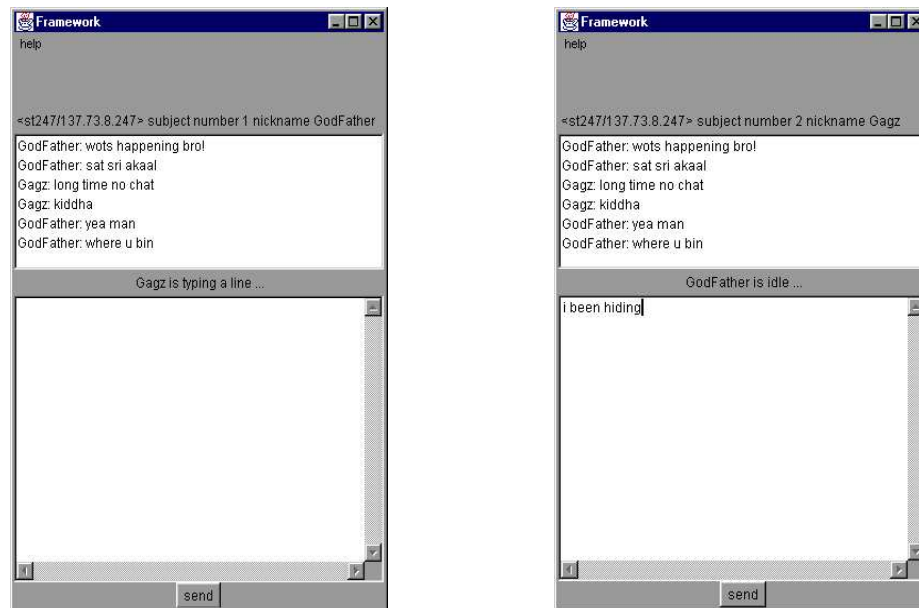
Figure 3.13: Chattool Client Interface

they are generated by the other participant. The client also has the ability to display an error message and prevent text entry: this can be used to delay one participant while the other is engaged in an artificially-generated turn sequence.

Server

Each turn is submitted to a server (also written in Java by King) on a separate machine. This server passes the text to a NLP module for processing and possible transformation, and then displays the original version to the originator client, and the processed (or artificially generated) version to the other client. The server records all turns, together with each key press from both clients, for later analysis. This data is also used on the fly to control the speed and capitalisation of artificially generated turns, to be as realistic a simulation of the relevant subject as possible.

NLP Module

The NLP component consists of a Perl text-processing module which communicates with various external NLP modules as required: PoS tagging can be performed using LTPOS (Mikheev, 1997), word rarity/frequency tagging using a custom tagger based on the BNC (Kilgarriff, 1997), and synonym generation using WordNet (Fellbaum, 1998).

Experimental parameters are specified as a set of rules which are applied to each word in turn. Pre-conditions for the application of the rule can be specified in terms of PoS, word

frequency and the word itself, together with contextual factors such as the number of turns since the last artificial turn was generated, and a probability threshold to prevent behaviour appearing too regular. The effect of the rule can be to transform the word in question (by substitution with another word, a synonym or a randomly generated non-word, or by letter order scrambling) or, as in this experiment, to trigger an artificially generated turn or sequence of turns (currently a reprise fragment, followed by an acknowledgement, although other turn types are possible).

The setup for the experiment described here consists of rules which generate pairs of EFs and subsequent acknowledgements[79], for proper nouns, common nouns, verbs, determiners and prepositions, with probabilities determined during a pilot experiment to give reasonable numbers of EFs per subject. No use is currently made of word rarity or synonyms.

The turn sequences are carried out by (a) presenting the artificially-generated EF to the relevant client only; (b) waiting for a response from that client, preventing the other client from getting too far ahead by locking the interface if necessary; (c) presenting an acknowledgement to that response; and (d) presenting any text typed by the other client during the sequence. This sequence did not always work perfectly, especially when CRs were not answered, but subjects always managed to repair any problems quickly – see figure 3.14 for an example. No subjects reported problems or unnaturalness with the dialogue.

---

[79]Acknowledgements are randomly chosen amongst: "ah", "oh", "oh ok", "right", "oh right", "uh huh", "i see", "sure".

Subject A's View		Subject B's View
A: thats your assumption, i am assuming he has brought them with hime	$\longrightarrow$	A: thats your assumption, i am assuming he has brought them with hime
A: sorry him	$\longrightarrow$	A: sorry him
B: brought them for what?	$\longleftarrow$	B: brought them for what?
	Probe $\rightarrow$	A: brought?
	Block $\leftarrow$	B: you said it
	Ack $\rightarrow$	A: i see
B: say agin i'm lost	$\longleftarrow$	B: say agin i'm lost
A: he can adminster morphine to susie, so she feels less pain?	$\longrightarrow$	A: he can adminster morphine to susie, so she feels less pain?
B: thats bollocks	$\longleftarrow$	B: thats bollocks
B: just chuck him out	$\longleftarrow$	B: just chuck him out
B: administer morphine?	$\longleftarrow$	B: administer morphine?
	Probe $\rightarrow$	A: morphine?
	Block $\leftarrow$	B: shes 7 months pregnent
	Ack $\rightarrow$	A: oh
B: not nine	$\longleftarrow$	B: not nine
B: you didnt read it did you?	$\longleftarrow$	B: you didnt read it did you?
A: as he is on the brink of discovering the cure, it must mean he is still working on the cure	$\longrightarrow$	A: as he is on the brink of discovering the cure, it must mean he is still working on the cure
A: did not read what?	$\longrightarrow$	A: did not read what?
B: dont worry	$\longleftarrow$	B: dont worry

Figure 3.14: Balloon Task Excerpt, Subjects 3 & 4

3.5.2 Procedure

Experimental Procedure

28 subjects were recruited, 20 male and 8 female, average age 19 years, from computer sci-
ence and IT undergraduate students. They were recruited in pairs, where the members of a
pair were familiar with one another and both had experience with some form of text chat.

Two tasks were used to elicit dialogue, a balloon debate and a story-telling task (the latter
following Bavelas et al., 1992). In the balloon debate subjects are presented with a fictional
scenario in which a balloon with three named passengers is losing altitude and about to crash:
they must discuss which of the passengers should jump and try to come to an agreement. This
involves repeated references to particular named individuals, which allows EFs of grounded
(second-mention) names to be tested. In the story-telling task subjects are asked to relate
any personal 'near-miss' story. This was chosen due to its unrestricted nature, to rule out
any effects of the restricted domain of the balloon task. To ensure that subjects concentrated
on understanding their partners' stories, they were asked (in advance) to write a summary
afterwards: this was intended to encourage CRs to be answered rather than ignored.

The chattool software was set up to generate EFs based on the PoS category and the
first/second mention of words in the input produced by the subjects. EFs were generated
for both first and second mention of proper names, common nouns, verbs, determiners and
prepositions. A maximum limit of 1 of each type per subject per conversation was imposed,
together with an enforced gap of at least 5 turns between each artificial CR. This generated a
total of 215 EFs.

Processing of Results

After the experiment, the logged dialogues were marked up in the same way as the previous
three corpus studies: for CR form, reading and source distance; for answer type and distance;
and for source PoS category and first/second mention.

PoS Category/Mention   The PoS category and the first/second mention nature were marked
automatically by the chattool software, but were still checked and corrected manually where
necessary. PoS categories were then grouped as `cont` for content words (verbs, common
nouns, proper names and pronouns), `func` for function words (determiners, prepositions,
conjunctions and complementisers) and `oth` for all others (greetings, interjections, "smileys",
chat-specific conventions such as "lol" etc.).

PoS required correction as the internal PoS tagger used by the chattool produced incorrect
word categories in approximately 30% of cases. This error rate may seem high, but the tagger
was trained on standard newspaper text, rather than "chat", which contains not only typing
errors, but chat-specific conventions (e.g. "k" for "okay"). Detection and classification of

proper nouns was also sensitive to capitalisation. Subjects were not consistent or conventional in their capitalisation of words and this caused some misclassifications.

First/second mention required correction as the chattool has no memory between conversations (CRs classified as first mention had to be checked to ensure that they hadn't already occurred in a previous dialogue) and contained a bug that prevented repeats with different case being recognised as such.

CR Form   As all examples generated were fragments echoed from a previous sentence, the purpose of markup for CR form was only to distinguish cases where the artificially generated turn was interpreted as a reprise *gap* rather than *fragment*: see figure 3.15. As the dialogue was text-based, no intonational cues were available to the subjects to prevent this ambiguity.

Subject A's View		Subject B's View

B: `we played on the`
   `house/market rooftop`
   `and`
	$\longleftarrow$	B: `we played on the`
		   `house/market rooftop`
		   `and`

	Probe $\rightarrow$	A: `on?`

	Block $\leftarrow$	B: `this one saturday we`
		   `played on the market`
		   `roof top`

	Ack $\rightarrow$	A: `ah`

B: `and we were playing`
   `Polo`
	$\longleftarrow$	B: `and we were playing`
		   `Polo`

Figure 3.15: Probe Turn treated as Gap, Subjects 19 & 20

The only categories used from the form classification of section 3.2 were therefore the `frg` and `gap` classes. One extra category was also added: non-clarificational (`ncl`), referring to cases in which the fragment was treated by the experimental subjects as something other than a CR (this did not apply when building the original corpus, as only utterances treated as CRs were considered).[80] Examples are shown in figure 3.16, where *"OK?"* is treated as meaning something like *"Are you ready?"*, rather than *"Did you say/mean 'OK'?"*; and figure 3.17, where *"were?"* appears to have been interpreted as the direct sluice question *"Where?"*.

---

[80]Note that this *non-clarificational* class `ncl` is not the same as the *non-reprise* class `non` in the corpus study: the latter were CRs, but expressed without using a reprise form.

Subject A's View		Subject B's View
A: who	$\longrightarrow$	A: who
B: erm	$\longleftarrow$	B: erm
A: hurry up	$\longrightarrow$	A: hurry up
B: ok	$\longleftarrow$	B: ok
B: ok	$\longleftarrow$	B: ok
	Probe $\rightarrow$	A: ok?
	Block $\leftarrow$	B: Mr tom
	Ack $\rightarrow$	A: uh huh
B: ?	$\longleftarrow$	B: ?
A: no pressure man, but hurry up	$\longrightarrow$	A: no pressure man, but hurry up

Figure 3.16: Probe Turn treated as Non-Clarificational, Subjects 9 & 10

Subject A's View		Subject B's View
B: then when we were finally done	$\longleftarrow$	B: then when we were finally done
	Probe $\rightarrow$	A: were?
	Block $\leftarrow$	B: that was when we lived in wembley
	Ack $\rightarrow$	A: right
A: ok	$\longrightarrow$	A: ok

Figure 3.17: Probe Turn treated as Non-Clarificational, Subjects 21 & 22

Response Type   This was classified using the same scheme as section 3.4, except that two new classes were added. Firstly, the `qury` class of responses (in which CRs are explicitly queried rather than answered) was also extended to include not only cases in which the CR is queried (figure 3.18), but also those in which the CR is recognised as such but the addressee refuses to answer (figure 3.19), or a combination of the two (figure 3.20).

Secondly, a new category `end` was introduced to account for cases where the time limit was reached and the experiment ended when a CR had just been generated but before a response could happen.

Note that the `uncl` class does not apply in this case (as this only referred to the `<unclear>` transcription in the BNC); nor does the `cont` class, as the apparent speaker (the chattool software) does not continue until a response (or at least a turn of some kind) is given.

Subject A's View		Subject B's View
B: firstly women cant drive how they gonna control a balloon and the doc needs to use his memory for scientific research not rubbish like this	⟵	B: firstly women cant drive how they gonna control a balloon and the doc needs to use his memory for scientific research not rubbish like this
	Probe →	A: this?
	Block ←	B: u what
	Ack →	A: oh
B: so i think it should be the woman	⟵	B: so i think it should be the woman

Figure 3.18: Explicitly Queried Clarification, Subjects 1 & 2

Subject A's View		Subject B's View
A: Not all French classes are boring	⟶	A: Not all French classes are boring
A: James contradictd	⟶	A: James contradictd
B: James?	← Probe	
A: Just get on wit it	→ Block	
B: right	← Ack	
A: tell u later	⟶	A: tell u later

Figure 3.19: Refused Clarification, Subjects 7 & 8

Form/Reading/Response Interdependence   In these experiments, as the CR was generated artificially rather than by one of the participants, then only way of judging the form and

Subject A's View		Subject B's View
B: which moron decided to throw the prachute out?	⟵	B: which moron decided to throw the prachute out?
B: i mean thats just kamakaze style?	⟵	B: i mean thats just kamakaze style?
A: jokes	⟶	A: jokes
A: i dont think they take sparachutes up in a hot air baloon	⟶	A: i dont think they take sparachutes up in a hot air baloon
B: take?	← Probe	
A: whats with you and these stupid one word questions	→ Block	
B: oh ok	← Ack	
A: wtf do u mean 'take?'	⟶	A: wtf do u mean 'take?'
B: its not jokes; its stupidity....	⟵	B: its not jokes; its stupidity....

Figure 3.20: Queried/Refused Clarification, Subjects 11 & 12

reading ascribed to it by the addressee is through their response. There is no relation to previous dialogue context except for the words in the previous sentence; we cannot relate the CR to the apparent speaker's goals or previous moves, as they were not really the speaker at all; and of course the apparent speaker will not pursue the question if it is not answered. In cases where no explicit response was forthcoming, it was therefore not possible to classify the artificial CR for reading, and essentially meaningless to classify it for form. Values for form and reading are therefore only given for those CRs that were responded to.

3.5.3 Results

Raw results are shown in table 3.33 for response type over all examples, and in table 3.34 for CR form and reading for those examples that did receive a response.

Response Type

50% of cases (109 of the 215 total) received no explicit response. This is even more than in section 3.3's corpus study, in which between 19% and 43% of EFs went unanswered (depending on whether we exclude or include possible unclear answers and cases where the speaker keeps the turn). This is not what was expected: use of text-based dialogue was intended to increase the likelihood of responses by ruling out non-verbal communication, and use of a debate in one task and the requirement to summarise afterwards in the other were intended to

	none	qury	end	frg	sent	yn	oth	*Total*
cont/1	29	2	1	12	8	17	1	70
cont/2	43	4	1	7	4	16	0	75
func/1	6	3	0	2	0	1	0	12
func/2	20	6	1	1	2	0	0	30
oth/1	6	0	0	4	0	3	0	13
oth/2	5	0	0	4	3	2	1	15
Total	109	15	3	30	17	39	2	215

Table 3.33: Response Type vs. PoS Category/Mention

increase the pressure to answer a CR.

This is likely to be due at least in part to the medium. Firstly, as participants only see each other's turns once they are finished (typing completed and RETURN key pressed), they often produce their turns simultaneously. This can result in a new turn arriving when a subject is half-way through typing a long sentence. They must then trade off the cost of undoing this turn to respond to the new one, against going ahead anyway and ignoring the new turn (possibly responding later if it seems necessary). Secondly, it is easy for a subject not to notice when a new turn arrives, especially if they are involved in typing. It may be possible to improve the design to reduce these effects, transmitting turns word-by-word or character-by-character, and adding an alert when new turns arrive. Still, both corpus and experimental work seem to indicate that a significant proportion of CRs may simply be ignored.

Form/Reading

Where responses did occur and form and reading could be classified, gap and non-clarificational cases were rare (4 and 9 instances respectively), with all other instances classed as reprise fragments. This overall frequency of gaps vs. fragments is similar to that found in section 3.2's corpus study.

	frg/cla	frg/con	frg/lex	gap	ncl	*Total*
cont/1	23	14	0	1	0	38
cont/2	16	7	0	0	4	27
func/1	0	0	1	2	0	3
func/2	0	0	0	1	2	3
oth/1	3	4	0	0	0	7
oth/2	2	5	0	0	3	10
Total	44	30	1	4	9	88

Table 3.34: Form/Reading vs. PoS Category/Mention

No correction readings were noted, and only one case classed as a lexical reading. No

lexical readings were expected in this text-based dialogue: the example is shown below in figure 3.21 – subject A appears to have interpreted the echo as a pathological question about word identity, or perhaps a play on words (this seems to illustrate how difficult echoes of some determiners can be to interpret as reprises). It is perhaps debatable whether this could in fact be classed as a clausal reading, but the meaning seems to concern the identity of the word used. In any case, all other cases had clausal or constituent readings, as expected.

Subject A's View		Subject B's View
A: `I havelimited experience with balloons`	$\longrightarrow$	A: `I havelimited experience with balloons`
A: `but...`	$\longrightarrow$	A: `but...`
A: `worth a try`	$\longrightarrow$	A: `worth a try`
B: `a?`	$\leftarrow$ Probe	
A: `no, b`	$\rightarrow$ Block	
B: `oh ok`	$\leftarrow$ Ack	
B: `i'm not in the baloon`	$\longleftarrow$	B: `i'm not in the baloon`

Figure 3.21: Balloon Task Excerpt, Determiner Clarification, Subjects 7 & 8

As before, the clausal reading is preferred over the constituent reading in general, with 65% of content-word `frg` cases taking this reading (59% over all PoS categories). Note however that this is not as strong a bias as seen in the corpus study.

Effect of PoS Category

PoS category has a large effect on likelihood of response: 65 of 143 content words (45%) are explicitly answered (responded to in a way other than refusal or querying the question)[81], but only 6 of 41 function words (15%). This difference is statistically significant: a $\chi^2_{(1)}$ test gives $p = 0.04\%$.

The effect on the attributed form and reading is also strong. No function word EFs seemed to be interpreted as fragments with clausal or constituent readings: all 6 which received explicit answers were classified as gaps or non-clarificational questions, or in the one case discussed above, a reprise fragment with a possible lexical reading. In contrast, 60 of 65 (92%) of answered content word EFs were interpreted as clausal or constituent reprise fragments, with only one being interpreted as a gap. This is as predicted, and is also a reliable difference: a $\chi^2_{(1)}$ test gives $p < 0.01\%$.

Interestingly, though, examining the PoS category in more detail (see tables 3.35 and 3.36) shows that there is no significant difference between nouns and verbs on the distribution of attributed form and reading, or indeed on the likelihood of response. Although section 3.3

---

[81]Examples classed as `end` are excluded from this calculation.

showed that verbs are unlikely to be the source of CRs, it seems that this is not because verb reprise forms (or at least reprise fragments) are hard to interpret as such, or hard to answer. We will have to look elsewhere for the reasons behind the apparent difference in likelihood of clarification – as suggested before, perhaps to semantics or information content.

	none	qury	end	frg	sent	yn	oth	*Total*
cn	32	0	2	9	4	10	1	58
pn	10	3	0	5	2	13	0	33
pro	6	2	0	1	1	3	0	13
Total (noun)	*66*	*4*	*2*	*18*	*11*	*30*	*1*	*132*
v	24	1	0	4	5	7	0	41
Total (verb)	*24*	*1*	*0*	*4*	*5*	*7*	*0*	*41*
Total	*72*	*6*	*2*	*19*	*12*	*33*	*1*	*145*

Table 3.35: Response Type vs. PoS Category (noun/verb only)

	frg/cla	frg/con	frg/lex	gap	ncl	*Total*
cn	15	9	0	0	0	24
pn	13	7	0	0	1	21
pro	3	1	0	0	1	5
Total (noun)	*31*	*17*	*0*	*0*	*2*	*50*
v	8	4	0	1	2	15
Total (verb)	*8*	*4*	*0*	*1*	*2*	*15*
Total	*39*	*21*	*0*	*1*	*4*	*65*

Table 3.36: Form/Reading vs. PoS Category (noun/verb only)

Effect of Grounding

Previous mention also has an effect on the likelihood of a response, although less marked: taking both content and function word EFs together, 41 of 81 (51%) of first mentions are explicitly answered, whereas only 30 of 103 (29%) of second mentions did ($p = 0.3\%$ according to a $\chi^2_{(1)}$ test). Including words in the other category makes this effect slightly less strong and less reliable (many of these words are "chat" expressions, whose behaviour we might expect to be different), but still significant: 51% of first mentions are answered vs. 34% of second mentions, with a $\chi^2_{(1)}$ test giving $p = 1.2\%$.

However, previous mention does not appear to have as strong an effect on the attributed form and reading as might be expected. Taken over all PoS categories, it does have an effect on the likelihood of a clausal or constituent reading, as opposed to being treated as a gap or non-clarification: only 4 of 48 first mentions (8%) are classified as gap, ncl or lex, compared to 10 of 40 second mentions (25%). This effect is statistically reliable ($\chi^2_{(1)}$ gives

$p = 3.3\%$), but perhaps not strong enough to be useful in disambiguation (at least on its own). When looking at content words only, there is a similar bias, but this is not reliable.

Similarly, while there does seem to be an effect on the likelihood of clausal vs. constituent reading (for content words, first mentions are more likely to be interpreted as constituent readings than clausal, as expected), the effect is not especially strong, and not reliable. For proper names, the effect does seem strong (clausal/constituent proportion of first mentions is 54%/46%, for second mentions 86%/14%), but is again not statistically reliable ($p = 15\%$) due to the small number of data points in this class.

This suggests that while level of grounding does have an effect on the reading, it is not as simple as looking at first or second mention. This makes sense, at least for common nouns and verbs, for which effects like word rarity or context might be more important (a very common word in a particular context is not likely to trigger a constituent CR even on first mention, as its meaning is essentially already mutual knowledge). For proper names first/second mention might be more accurate as a method of disambiguation (and the correlation did appear stronger), although domain, context and shared history will presumably also play a part.

3.5.4 Conclusions

The main conclusions we draw from the experimental results presented in this section are as follows:

- Reprise CRs appear to go without response far more often than might be expected, both in the BNC and in our experimental corpus. Both may be effects of the media (transcription in one case, turn sequencing overlap in the other), and the effect may also change in more critical task-oriented domains, but the figures are large enough and similar enough to warrant further investigation.

- Word PoS category seems to be a reliable indicator of CR form: echoed function words are likely to be reprise gaps (where the original function word is only a secondary source), rather than reprise fragments. This can help us in disambiguating user CRs, and in choosing forms when generating system CRs.

- EFs generated on the first mention of a word have a higher likelihood of receiving a response, and of being interpreted as a reprise fragment, than on second mention. While first mention may also make constituent readings more likely than clausal readings (especially for proper names), we cannot be sure.

It is also worth noting that the new technique introduced here seems viable: the chat-tool was successful in producing plausible clarification sequences — although in some cases participants had difficulty making sense of the artificial clarifications this did not make them distinguishable from other, real, but equally problematic turns from other participants — and

the fine-grained manipulations that it allows to be directly introduced into dialogues make it a powerful tool that may lead to further useful investigations in the future.

3.6   Summary

The corpus studies and experiments presented here have gone a long way towards answering the questions posed at the beginning of the chapter.

What forms do CRs take, and what readings can these forms have?

Section 3.2 has presented the possible readings and forms derived from a sub-part of the BNC. All of these readings and forms seemed to lend themselves to G&C's analysis or to analyses along the same lines (i.e. based on grammar and limited contextual operations rather than general inference). As far as correlations between forms and readings go, there are some strong effects, with the common forms all tending towards a particular reading (literal reprises, sluices and fragments being clausal, conventional CRs and fillers being lexical).

How common are CRs (and the various forms they can take)?

CRs in general are common, making up 3-4% of turns in dialogue. The most common forms are the conventional and reprise fragment forms, with non-reprise CRs and reprise sluices coming next.

When do CRs occur (how long after the phrase being clarified)?

About 80% of CRs occur on the very next turn, and almost all occur within 4 turns: we can assume that this length of utterance memory is enough for clarification purposes within a dialogue system.

How do form and reading depend on the type of phrase being clarified?

It seems that function words are unlikely to get clarified at all (and hence that echoed function words are more likely to be reprise gaps), at least in the rather general dialogue domain examined here. Verbs also seem only rarely to be sources of CRs, with most CRs asking about whole utterances (in which case, conventional CRs are used) or sentences (literal reprises), or about NPs (most other forms). Expanding the grammatical analysis to various kinds of nouns and NPs therefore seems vital, but other classes of word and phrase are less important.

How and when are CRs answered?

Many CRs were not actually answered at all, although non-reprise CRs, and CRs asked on the first mention of the source, seem to be more likely to get responses than others. Those

that were answered were usually answered in the very next turn. The nature of the answer can often be determined from the form of the CR (conventional CRs are answered with full sentences, sluices with fragments, and literal reprises with yes/no answers), but in the case of reprise fragments, the usual answer type depends on reading (fragment for constituent, yes/no for clausal). Assigning reading correctly (based on the observed effects of source category, and the possible effects of level of grounding) will therefore be very important.

Chapter 4

Implications for Semantics

4.1 Introduction

We have seen that CRs can request clarification of the meaning intended by a speaker when uttering a word or phrase. In the case of a proper name, as shown in the original G&C examples in section 2.3.5, the intended meaning is presumably its referent, the entity which bears that name. But what of other word and phrase types? This chapter attempts to answer the question of what CRs really query when clarifying the meaning of other phrase types, and therefore both shed some light on what a sensible semantic representation for these phrase types might be, and thereby also allow a grammatical analysis of CRs to be extended to cover them.

4.1.1 Overview

In this chapter we are concerned specifically with those CRs which concern the *meaning* intended by a speaker when uttering a word or phrase: in other words, clausal and constituent questions (which we shall refer to as *content* questions), but not lexical ones. By examining what these questions appear to be asking about, they can provide us with information about what meaning can be associated with word and phrase types. In other words, we can use them as *semantic probes*, and thereby provide useful extra evidence for the field of semantics – a domain overfull with theories underdetermined by evidence. We are therefore interested in those CRs which (a) tend to ask content questions rather than lexical ones, and (b) specify their source accurately, allowing us to be sure which original word or phrase is being asked about. This means that we will mainly examine reprise fragments, although reprise sluices will also be considered when useful.

This chapter discusses the evidence provided by these reprise questions concerning the semantics of, firstly, the phrase types which seem to be the most common sources of clarification — common nouns (CNs), pronouns and quantified noun phrases (QNPs) — and then,

more briefly, determiners, verbs and verb phrases. It extends G&C's analysis to cover these, and outlines some general implications for NP semantics, together with some implications for semantic representation in HPSG and other related underspecified representations.

The central finding is that while the traditional views of nouns and verbs as properties of individuals seem tenable, reprise questions strongly suggest that QNPs denote (situation-dependent) individuals — or sets of individuals — rather than sets of sets, or properties of properties, as might be expected given a traditional approach to semantics. This leads to a witness-set-based analysis which treats all QNPs in a coherent manner, and allows an analysis of reprise questions which follows the approach outlined so far. It also shows how anaphora and quantifier scope can be accounted for within this analysis, via a view of NPs as functional, and shows how non-monotone-increasing NPs can be represented.[1]

4.1.2   Reprises as Semantic Probes

G&C's analysis of proper name (PN) reprise fragments treats them as questions concerning the entire semantic content of the PN (which is taken to be a parameter with a referential index, the intended referent of the name). In this way, a reprise such as that in example (117) can be taken to be paraphrasable as shown:

(117)

	A:	Did Bo leave?
	B:	**BO?**
	$\rightsquigarrow$	*"Is it **BO**$_i$ that you are asking whether i left?"*
	$\rightsquigarrow$	*"Who do you mean by 'Bo'?"*

The two paraphrases correspond to the distinct clausal and constituent readings, but both concern the content of the PN *Bo*, via the now familiar analysis of section 2.3.5: the PN includes its content, a referential parameter, as a member of its contextually dependent C-PARAMS set. This is inherited up to the sentence level, making the whole utterance contextually dependent in that this parameter must be instantiated in context (grounded), and failure to do so can result in a CR concerning the problematic parameter. The evidence seen so far seems to support this analysis: all the reprise questions found do appear to ask about intended content (or of course, in the case of lexical readings, surface form); and echoed function words (which one would expect to have little or no contextually dependent content) seem very difficult to interpret as reprises.

G&C's analysis applies only to PNs. However, it is clear that other fragments can be reprised, and the intention of this chapter is to examine such reprises and, where possible, propose a suitable extended analysis. As chapter 3 has shown, it is most important to provide this for NPs of various types. It is also clear that not all reprises involve querying a simple referential index: exactly what a reprise question can query is likely to vary depending on

---

[1]Much of the material in this chapter has appeared as (Purver and Ginzburg, 2003, 2004).

the nature of the source fragment itself. On the other hand, if content readings of reprises are questions about part of the semantic content of the source fragment, then examining them can give some evidence about what goes to make up that semantic content.

Do Reprises Query Content?

One can imagine an argument that reprises can query any aspect associated with meaning, including perhaps pragmatic inferences, and that it might therefore be difficult to tease apart semantic from pragmatic readings. However, there is good reason to believe that while some CRs may be able to query some material of a pragmatic nature, queries about inferences in general (including implicatures and the like) are very difficult if not impossible to construct.

Pragmatic Readings   Some CRs do seem to be able to query the whole intended speaker's meaning, or even the relevance of the utterance to the discourse. Many of those conventional CRs which were marked as having a constituent reading in the previous section could be taken to be in this class. An example is shown here in example (118) – here the question asked seems more about the utterance's relevance or overall intended meaning, rather than its semantic content or predicate-argument structure:

$(118)^2$

Sheila:	. . . when Michael's in she knits him a jumper, the jumper `<unclear>` `<pause>`
Wendy:	Best that way then you don't get sick
Sheila:	**Eh?**
Wendy:	It'll be better that way if you, like you're knitting with two different colours
Sheila:	Aye

$\rightsquigarrow$ *"What do you mean by 'Best that way then you don't get sick'?"*

It seems much more difficult for reprise CRs to ask this sort of question, but although none were found, it can be imagined for reprises of whole sentences – a reprise *"Best that way then you don't get sick?"* in example (118) might serve the same purpose as the conventional CR. It seems much harder to imagine sub-utterance fragment reprises asking these sort of questions, though, and the corpus work did not reveal any, even though reprise fragments are very common.

Inferences   However, this is a far cry from being able to query inferred pragmatic meaning in general. Examples involving implicatures suggest that it is very difficult for reprise questions to query pragmatically inferred content. It is certainly the case that A's statement in the invented example (119), taken to be uttered outside a West End theatre currently showing a

---

[2]BNC file KR0, sentences 362–366

best-selling musical, could be inferred to be implicating other messages as shown:

(119)
	A:	I have a ticket for tonight's performance.
	$\leadsto$	*"I am offering to sell a ticket for tonight's performance."*
	$\leadsto$	*"Would you like to buy a ticket for tonight's performance?"*

But a reprise of the sentence does not seem to be able to be understood as querying these implicatures, but only the *directly conveyed semantic content*, as shown in example (120).

(120)
	A:	I have a ticket for tonight's performance.
	B:	**You have a ticket for tonight's performance?**
	$\leadsto$	*"Are you telling me you have **a ticket**?"*
	$\leadsto$	*"What do you mean by 'you have a ticket'?"*
	$\leadsto$	#*"Are you offering to sell me a ticket?"*
	$\leadsto$	#*"Are you asking if I want to buy a ticket?"*

This may be even clearer when considering an answer to such a reprise question (example (121)), which again can only be construed as answering a question about this directly conveyed content (see Ginzburg et al., 2001b, 2003, for a more detailed exposition):

(121)
	A:	I have a ticket for tonight's performance.
	B:	**You have a ticket for tonight's performance?**
	A:	Yes.
	$\leadsto$	*"Yes, I am indeed telling you I have a ticket."*
	$\leadsto$	#*"Yes, I am indeed offering to sell you a ticket."*
	$\leadsto$	#*"Yes, I am indeed asking if you want to buy a ticket."*

Note that *"Yes, but I'm not offering to sell it"* would be perfectly acceptable. Similarly *"No"* must mean *"No, I do not have a ticket"*, rather than *"No, I'm not offering to sell a ticket (although I might have one)"*. Any inference that B is really asking about buying or selling activities therefore seems to be exactly that – an inference on top of the content of the reprise (a question about content of the original utterance), rather than because the reprise is itself a question about inferred material.

It should also be noted that the empirical results from both corpus and experimental work in chapter 3 showed that function words were very unlikely to be the source of CRs, and that function word questions were extremely difficult to interpret as reprises. If reprises ask about semantic content, and in particular some sort of contextually dependent reference, this makes sense. On the other hand, if they could be based on unrestricted contextual inferences, one might expect that such inferences (and resulting reprise readings) would be available even for function words.

So while some reprises can be seen as querying pragmatic material (such as overall relevance), they do not appear able to ask about unrestricted pragmatic inferences, and in most cases really do seem to query semantic content (particularly when querying fragments rather than whole utterances). We can therefore take it that fragment reprises which appear to query semantic content (rather than, say, phonology) really are doing so.

Strengthening Compositionality

Given this, it seems clear that if a question which reprises a particular source phrase asks about a particular semantic object, then that object must be part of the semantic representation of the source phrase. In other words, reprise questions must query at least *some part of the semantic content* of the fragment being reprised, and we take this as our basic hypothesis:

Reprise Content Hypothesis (weak version):

(122) *A reprise fragment question queries a part of the standard semantic content of the fragment being reprised.*

A stronger proposal might be that if a reprise question asks about a particular semantic object, then that object *is* the semantic content of the phrase being reprised:

Reprise Content Hypothesis (strong version):

(123) *A reprise fragment question queries exactly the standard semantic content of the fragment being reprised.*

While there is (and can be) no independent evidence that this stronger version holds, it is intuitively very attractive, as it provides a version of Occam's Razor: it requires that we do not postulate any part of a semantic representation which cannot be observed via a reprise question – in other words, that the semantic representations we do postulate are the simplest possible that can explain the readings of reprise questions.

This hypothesis, in either version, provides an empirical criterion for assigning denotations that supports, but is stronger than, the usual criterion of compositionality. The standard requirement that the full content of an utterance (or sentence) emerges from the contents of its components often leaves underdetermined the question of which part contributes what. Instead, as originally observed by Ginzburg (1999), a semantics that can provide an adequate analysis of reprise questions by holding to the reprise content hypothesis is held responsible for the content it assigns not only to the complete utterance but to each component (or at least each reprisable and semantically potent component). A suitable semantics for NPs must not only allow full sentence content to be built, but be able to explain what it is about NPs that gives NP reprises the meanings that they appear to have.

Throughout the chapter, then, we will examine the consequences of both versions of this hypothesis for NP semantics, proposing representations which always hold to the weak version, and hold to the strong version wherever possible.

4.1.3   Corpus Evidence

As before, the BNC has been used to provide empirical data – to provide actual occurrences of reprise questions in dialogue. This time, the whole BNC was used, rather than a sub-corpus, and questions were found automatically using SCoRE (Purver, 2001), by searching for common reprise patterns (e.g. words repeated from the immediately preceding turn for fragments, bare *wh*-words for sluices). This method means that some examples may have been missed, but provides us with a lower bound: at least those questions that were found must be accounted for by a semantic theory.

The resulting examples were then classified according to possible and impossible para-phrases – these are of course constructed subjectively, but every effort has been made to infer them not only from the questions themselves but from the dialogue context, particularly the responses given by the other participants in the dialogue. Possible paraphrases are therefore those which seem consistent both with the question and the recorded responses, and impossible ones those which would be inconsistent with either. This process has not been repeated by another independent marker, but the similarity to the markup process of chapter 3 (which gave good statistical reliability) gives some hope that it would be repeatable.

The primary purpose in using a corpus in this chapter is to provide as many examples as possible, in different situations, with different words and phrases (tokens as well as types) and with different speakers, in order to give some confidence that any claims about possible question readings are not influenced by subjective choice of imagined examples. While CRs in general are common, the reprise forms we are interested in here, or more accurately those examples of these forms that fit the patterns which we are able to search for, are rare enough to make data sparsity an issue. As it turns out, the BNC is large enough (the dialogue portion comprises 740,000 speaker turns) to provide a few dozen occurrences for each of the phrase types we are most interested in here – that is, CNs and definite & indefinite NPs (exact num-bers are given in the relevant sections below). While this quantity of data is small compared to the samples usually used for statistical studies, it certainly fulfils the primary purpose by providing a significant number of examples that must be covered by any proposed analysis. It also provides enough data to ensure that the observed differences in reading distribution for these phrase types are statistically significant according to χ^2 tests, as detailed below.

However, even a corpus of this size yielded very few (< 10) examples of reprises of some other classes: in particular, NPs with quantifiers other than definite & indefinite determiners, and determiners themselves. In the corresponding sections the sample therefore has to be augmented using intuition and invented examples, but I have indicated below where this is the case, and have not attempted to draw any conclusions based on statistical distributions or apparent negative evidence, but only ensured that any observed examples are accounted for.

The next section gives some background on traditional views of the semantics of QNPs and verbs. The subsequent sections 4.3 and 4.4 discuss the content of reprise questions for

CNs and QNPs together with a corresponding semantic analysis, and some further issues arising from this are discussed in section 4.5. Section 4.6 then examines other word and phrase types.

4.2 Background: QNP Semantics

The semantic representation of QNPs has of course been a subject of lively debate for some time, and there is little point trying to do justice to the field here; instead this section points out the main differences in currently popular views in the areas on which the study of reprise questions may shed some light.

4.2.1 The Quantificational View

One view, dating back at least to Russell (1905), holds that QNPs contribute quantificational terms to the semantic representation of a sentence. This is exemplified by Montague (1974)'s PTQ, in which sentences containing QNPs are given representations as follows:

(124) *"every dog snores"* $\quad\mapsto\quad \forall x(dog(x) \rightarrow snore(x))$

On this view, QNPs therefore denote functions from properties of individuals ($e{\rightarrow}t$) to truth values (t) (in other words, they are properties of properties ($(e{\rightarrow}t){\rightarrow}t$)). The content of a QNP is defined by the properties that hold of some referent contained in it (in the case of *"every dog"*, all those properties which are true or untrue of every dog).

(125) *"every dog"* $\quad\mapsto\quad \lambda P.\forall x(dog(x) \rightarrow P(x))$

Those who adhere strictly to this view take it also to hold for definite descriptions: definites are not considered to be directly referential in the same sense as PNs, but are seen as defined by existential quantification with a uniqueness constraint.

(126) *"the dog"* $\quad\mapsto\quad \lambda P.\exists x(dog(x) \wedge \forall y(dog(y) \rightarrow y = x) \wedge P(x))$

4.2.2 The Referential View

An alternative view originating with Strawson (1950) and Donnellan (1966) is that some NPs, in particular definites, can be directly referential. Donnellan pointed out that while Russell's approach covered *attributive* uses well (those described by Russell as "known by description"), it did not appear to cover *referential* uses. Others (e.g. Fodor and Sag, 1982) have also pointed out that indefinites can be used *specifically* (the speaker has a specific individual in mind, although the hearer is not expected to be able to identify it) and *definitely* (expected to be identified by the hearer)[3], and that these uses also do not appear to fit with a purely

---

[3]A good summary of these terms, with examples, is available in (Ludlow and Neale, 1991).

quantificational analysis.

On the quantificational view, this apparently referential nature is argued to follow from pragmatic principles rather than any true semantic reference. This argument originates with Kripke (1977), and a concise statement is given by Ludlow and Neale (1991) and Ludlow and Segal (2004). Essentially it runs as follows (omitting some steps for brevity's sake here):

1. S has expressed a quantified proposition $\tau x.F(x) \wedge P(x)$.
2. S could not be doing this unless she thought that $P(b)$ where b is some referent.
3. S knows and I know that $b = \tau x.F(x)$
4. Therefore S has implicated that $P(b)$.

Other approaches such as the dynamic theories of Heim (1982), Kamp and Reyle (1993) and possibly Groenendijk and Stokhof (1991) might be said to fall somewhere in between the two camps, with definites having some kind of reference (although this may be to a contextual discourse referent rather than a real-world object).

In most views, however, NPs with other quantifiers (*every, most* etc.) are seen as quantificational.

4.2.3   Generalized Quantifiers and Witness Sets

The theory of Generalized Quantifiers (GQs) (see Barwise and Cooper, 1981)[4] (hereafter B&C) has been applied to the quantificational view, both to extend the Russellian approach to other natural language quantifiers, and to allow semantics of the QNP constituent to be represented more transparently in the sentence representation:

(127) *"every dog"* $\quad \mapsto \quad every(DOG) \quad$ where $\quad [\![every(DOG)]\!] = \{X | DOG \subseteq X\}$

(128) *"every dog snores"* $\quad \mapsto \quad every(DOG)(SNORE)$
where $\quad [\![every(DOG)(SNORE)]\!] = SNORE \in [\![every(DOG)]\!]$
$\qquad\qquad = SNORE \in \{X | DOG \subseteq X\} = DOG \subseteq SNORE$

Essentially the quantificational view of QNPs still holds: QNPs are GQs, and as such denote a family of sets (a set of sets, here the set of those sets which contain DOG, the set of dogs), rather than being directly referential.

To explain how a hearer can process a GQ without having to determine the identity of this full set of sets, B&C introduce the notion of a *witness set*. For a GQ $D(A)$, this is defined as being any set w which is both a subset of A and a member of $D(A)$. For an indefinite *a dog*, w can be any nonempty set of dogs; for a definite *the dog*, w must be the set containing exactly the contextually unique dog; for the universal *every dog*, w must be equal to the set of

---

[4]But see also e.g. (Keenan and Stavi, 1986; Keenan and Westerståhl, 1997; van der Does and van Eijck, 1996).

all dogs. For monotone increasing (MON↑) quantifiers, the following equivalence holds:

(129) $\exists w[w \subseteq X] \quad \leftrightarrow \quad X \in D(A)$

In other words, showing that a predicate X holds of a witness set is equivalent to showing that the corresponding GQ holds of the predicate. We will use this notion heavily below.

The next section begins by examining CN reprise questions, and shows that G&C's analysis can be extended to account for their apparent meaning in a manner consistent with traditional views of CN semantics. Section 4.4 then discusses QNP reprise questions, and shows that their meaning can be more naturally accounted for by the referential view of QNP semantics. Section 4.5 then discusses some issues raised by the view put forward in section 4.4, and section 4.6 them goes on to briefly examine some other phrase types.

4.3   Common Nouns

This section examines CN reprise questions, and shows that their meaning appears to be entirely consistent both with the standard semantic view of CNs as denoting properties of individuals, and with the hypothesis that reprise questions concern the semantic content of the fragment being reprised.

4.3.1   Nouns as Properties

The semantic content of CNs is traditionally viewed as being a property (of individuals). Montague (1974) expressed this as a λ-abstract, a function from individuals to truth values (e.g. $\lambda x.book(x)$), and this view is essentially shared by most strands of formal semantics. Variations (especially in representation) certainly exist: in situation semantics (Barwise and Perry, 1983) this might be expressed as a λ-abstracted infon (see Cooper, 1995), in DRT (Kamp and Reyle, 1993) as a predicative DRS (see Asher, 1993), but these approaches share the basic view that CNs are properties of individuals.

Given this, we would expect CN reprise questions to be able to query the property expressed by the noun, and this property only, when the hearer cannot identify this property in context. The clausal and constituent readings may both still be available, but the noun property or predicate should always be the element under question:

Clausal reading: *"Is it the property P about which you are asking/asserting . . . P . . . ?"*
Constituent reading: *"What property P do you intend to be conveyed by the word N?"*

In contrast, it should not be possible for CN-only reprises to be interpreted as questions about e.g. individual referents.

For mass nouns and bare plurals, the picture may not be so simple: these might be expected to refer instead to *kinds* (see e.g. Carlson, 1977; Chierchia, 1998), or in the case of

plurals, behave as indefinites (Kamp and Reyle, 1993) – or be ambiguous between the two (Wilkinson, 1991). Both are examined below in sections 4.3.4 and 4.3.5.

4.3.2 Corpus Evidence

Reprises of CNs were identified in the corpus by searching for single-word CN questions where the word is repeated verbatim from the previous speaker turn. To rule out bare mass nouns and plurals, which are discussed separately in sections 4.3.4 and 4.3.5, examples were restricted to cases in which the original occurrence of the CN in the previous turn was singular and preceded by a determiner. All examples found confirmed the expectation: as Table 4.1 shows, a predicate reading seems to be the only interpretation.

	Pattern	Referent Reading	Predicate Reading
CN Examples	*"... DET N ... " / "N?"*	-	58
		-	100%

Table 4.1: Literal Reprises – CNs

Examples are given here together with what appear to be possible and impossible paraphrases:

$(130)^5$

Monica:	You pikey! Typical!
Andy:	**Pikey?**
Nick:	Pikey!
Andy:	What's pikey? What does pikey mean?
Monica:	I dunno. Crusty.

$\rightsquigarrow$ *"Are you saying I am a **pikey**?"*
$\rightsquigarrow$ *"What property do you mean by the word 'pikey'?"*
$\rightsquigarrow$ #*"Which pikey are you saying I am?"*

---

[5] BNC file KPR, sentences 218–225

The same appears to be true when the CN reprised forms part of an indefinite NP:

(131)[6]

Emma:	Got a comb anywhere?
Helena:	**Comb?**
Emma:	Even if it's one of those `<pause>` tremmy [sic] pretend combs you get with a Barbie doll, oh this'll do! `<pause>` Don't know what it is, but it'll do!
$\rightsquigarrow$	*"Is it a **comb** that you are asking if I've got?"*
$\rightsquigarrow$	#*"Which comb are you are asking if I've got?"*

And indeed even when the CN is part of a seemingly referential definite NP:

(132)[7]

Carol:	We'll get the turkey out of the oven.
Emma:	**Turkey?**
Carol:	Well it's `<pause>` it's `<pause>` er `<pause>` what's his name? Bernard Matthews' turkey roast.
Emma:	Oh it's looks horrible!
$\rightsquigarrow$	*"Are you saying the thing we'll get out is a **turkey**?"*
$\rightsquigarrow$	*"What concept/property do you mean by 'turkey'?"*
$\rightsquigarrow$	#*"Which turkey are you saying we'll get out?"*
$\rightsquigarrow$	#*"Is it this/that turkey you're saying we'll get out?"*

Note that paraphrases which concern an intended referent of the NP containing the CN (e.g. the *"Which X . . . "* paraphrases) do not appear to be available, even when the NP might appear to be referential (see example (132)).

4.3.3 Analysis

As expected, we therefore suppose that the semantic representation of a CN must consist at least partially (and, if we are to hold to our strong hypothesis, solely) of a property of individuals.

An analysis entirely parallel to that of G&C is possible if properties of individuals (which we shall refer to here as *predicates*) are regarded as possible cognitive or contextual referents: that is to say, as entities that must be identified in context.[8] The predicate content of a noun can then be contextually abstracted by being made a member of C-PARAMS; this means it must be grounded by the hearer (by finding the intended predicate referent given its name) or made the subject of a clarification question in case this grounding process fails. It may fail for various reasons: with lexically ambiguous words, more than one property with this name will exist;

---

[6]BNC file KCE, sentences 1513–1516

[7]BNC file KBJ, sentences 131–135

[8]Whether these entities are best taken in a model-theoretic sense to denote atomic concepts (Barwise and Perry, 1983) or sets of individuals (Montague, 1974) is an interesting question in itself, but not one that impacts on the basic analysis here.

with unknown words, no known property may be found in context; in other cases the hearer may find the apparently intended predicate surprising or impossible. Noun content therefore becomes contextually dependent, rather than *a priori* given, as we require for a treatment of clarification. This may also offer a way to account for the psycholinguistically observable fact that conversational participants can have different understandings of the predicate being conveyed, and can indeed establish their own agreed meanings (see e.g. Pickering and Garrod, 2004).

A sensible representation of CNs therefore seems to be one in which the content (and the sole abstracted parameter in C-PARAMS) is a parameter whose INDEX value is a named property of individuals, as shown in AVM (133) both as an AVM and as an equivalent λ-abstract (see section 2.3.5):

$$(133)\quad \begin{bmatrix} \text{PHON} & \left\langle \text{dog} \right\rangle \\ \text{CONTENT} & \boxed{1}\left[P : name(P, dog)\right] \\ \text{C-PARAMS} & \left\{\boxed{1}\right\} \end{bmatrix} \qquad \lambda\left\{P\right\}[name(P, dog)].P$$

Comparison with Standard Approaches

This may seem uncontentious, but note that it does not correspond to the treatment of CNs by standard HPSG approaches to semantics. In the common unification-based approach (Sag and Wasow, 1999; Ginzburg and Sag, 2000), CN content is identified with that of the NP mother, and thus taken to be a parameter whose referent is an individual (the NP referent). Abstracting this parameter to C-PARAMS, as shown in AVM (134), would not give the correct reading for a clarification question, as this individual would become the referent to be grounded and thus the subject of the question (which we have seen is impossible).

$$(134)\quad \begin{bmatrix} \text{CONTENT} & \boxed{1}\left[x : dog(x)\right] \\ \text{C-PARAMS} & \left\{\boxed{1}\right\} \end{bmatrix} \qquad \lambda\left\{x\right\}[dog(x)].x$$

Avoiding this problem by abstracting only the relevant predicate rather than the entire content, as suggested in (Purver, 2002) and shown in AVM (135), would be possible but no longer holds to the strong hypothesis: as a result, clarification questions would not be able to query the entire semantic content, and we would be left with no explanation as to why not.

$$(135)\quad \begin{bmatrix} \text{CONTENT} & \left[x : P(x)\right] \\ \text{C-PARAMS} & \left\{\left[P : name(P, dog)\right]\right\} \end{bmatrix} \qquad \lambda\left\{P\right\}[name(P, dog)].x : P(x)$$

Similar problems apply to approaches such as Minimal Recursion Semantics (Copestake et al., 1999) in which the content of a NP mother is constructed by set union (*amalgamation*) over the content of its daughters (sets of *elementary predications*, simple pieces of propositional information). This again results in CN content including the individual referent of the

mother NP: making the entire content contextually available would seem to give the wrong readings for reprise questions; making only part of it available seems arbitrary.

$$(136) \quad \begin{bmatrix} \text{CONTENT} & \begin{bmatrix} \text{HOOK} \mid \text{INDEX} & x \\ \text{RELS} & \left\{ \begin{bmatrix} h1 : dog(x) \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix}$$

The predicate analysis proposed above seems preferable, as it holds to the strong hypothesis and thus explains why only the observed predicate reading of a reprise question is available. As discussed in section 4.4.5 below, this has implications for the usual inheritance and amalgamation principles used in HPSG.

4.3.4 Bare Singulars

As mentioned above, bare singular mass nouns might be expected to refer to kinds or concepts, but again not to individual referents. And again, this did appear to be the case. All reprises of bare singular CNs (i.e. singular CNs where the CN in the original utterance being clarified had no determiner) seemed to fit with this (see table 4.2).

(137)[9]

Richard:	because Donna is high in admir- admiration in fact I
Anon 4:	**Admiration**?
Richard:	I admire
Anon 4:	I think it's called infatuation
$\rightsquigarrow$	*"Is it the property/concept **admiration** you're saying Donna is high in?"*
$\rightsquigarrow$	*"What property/concept/kind do you mean by 'admiration'?"*

(138)[10]

Iris:	Oh you should see <pause> see it!
	<pause> It has only been <pause> burning coal in it!
Gordon:	**Coal?**
Iris:	And it's all burnt, it's burnt all the skirting board and er
Gordon:	Good God!
$\rightsquigarrow$	*"Is it the concept/kind/substance **coal** you're saying was burning?"*
$\rightsquigarrow$	*"What concept/kind/substance do you mean by 'coal'?"*
$\rightsquigarrow$	#*"Which individual bits of coal are you saying were burning?"*

Note that distinguishing between concepts, kinds and the properties or predicates discussed above has not been attempted, as this level of distinction does not seem possible from the imputed paraphrases – what is clear is that these sort of paraphrases always seem acceptable.

---

[9]BNC file KSV, sentences 5869–5874
[10]BNC file KCF, sentences 1573–1577

	Pattern	Referent Reading	Relation Reading	Predicate/Kind Reading
Bare Singular	*"...N..." / "N?"*	-	-	41
		-	-	100%
Bare Plural	*"...Ns..." / "Ns?"*	2	1	26
		3%	7%	90%

Table 4.2: Literal Reprises – Bare CNs

The analysis of mass nouns can therefore take exactly the same form as that for other CNs given above, with the semantic content being a property or kind which must be identified in context:

(139)
$$\begin{bmatrix} \text{PHON} & \langle \text{admiration} \rangle \\ \text{CONTENT} & \boxed{1}\big[P : name(P, admiration)\big] \\ \text{C-PARAMS} & \{\boxed{1}\} \end{bmatrix} \quad \lambda\{P\}[name(P, admiration)].P$$

4.3.5 Bare Plurals

With bare plurals, the situation was more complex. Most examples found did seem to follow the same lines, with a property or kind reading being preferred, and often being the only possible reading (see example (140)).

(140)[11]

John:	Now I would like you to tell me about numbers.
Simon:	**Numbers?**
John:	Mhm. What are they?
Simon:	**Numbers** `<laugh>` erm `<pause>`
John:	What do we use them for?

$\leadsto$ *"Is it things with the property **numbers** you're saying I should tell you about?"*

$\leadsto$ *"Is it the concept/kind **numbers** you're saying I should tell you about?"*

$\leadsto$ #*"Which numbers are you saying I should tell you about?"*

However, a few examples afforded a possible individual referent reading (with this seeming the preferred reading in two cases, examples (141) and (142)), and one example was best

---

[11]BNC file FMF, sentences 591–596

read as querying the plurality relation itself (example (143)).

$(141)^{12}$

Dorothy:	Anyway, you were telling me about `<pause>` meals.
Andrew:	**Meals?**
Dorothy:	Mm.
Andrew:	What `<unclear>`?
Dorothy:	At Pontepool.

$\rightsquigarrow$ *"Which meals are you saying I was telling you about?"*

$\rightsquigarrow$ *"Which property/concept do you mean by 'meals'?"*

$\rightsquigarrow$?*"Is it the property **meals** you're saying I was telling you about things with?"*

$(142)^{13}$

Rachel:	D'ya know what, I'm just gonna make up signatures.
	Cos I haven't asked anyone
Unknown:	**Signatures?**
Rachel:	I've taped.
Unknown:	Oh that's alright.

$\rightsquigarrow$ *"Which signatures are you saying you're going to make up?"*

$\rightsquigarrow$ *"Which property/concept do you mean by 'signatures'?"*

$\rightsquigarrow$ *"Is it **signatures** you're saying you're going to make up?"*

$(143)^{14}$

William:	You two
Unknown:	`<unclear>`
William:	hours ago
Clare:	`<laugh>` `<pause>` **Hours?**
William:	Well an hour
Unknown:	`<unclear>`
Kim:	it wasn't hours

$\rightsquigarrow$ *"Is it really **more than one** hour ago you're telling me it was?"*

As we will see in section 4.4.2 below, these are exactly the readings that seem to be available for indefinite NPs (a predicate reading, a logical determiner relation reading, and a (rarer) individual referent reading). This therefore suggests that bare plurals could be represented as indefinites (and we leave the details of this representation to section 4.4.2). However, as some examples seemed to *only* allow a property/kind reading (e.g. example (140) above), it may be that not *all* bare plurals are necessarily indefinites, but that (as assumed by Kamp and Reyle, 1993) they are best seen as ambiguous between indefinites and kinds.

---

[12]BNC file KBW, sentences 1247–1251

[13]BNC file KP5, sentences 1988–1992

[14]BNC file KBN, sentences 1367–1371

4.3.6 Summary

This section has presented evidence that shows that CN reprise questions concern a predicate. This seems consistent with the view shared by most semantic theories that the semantics of nouns are properties of individuals, and seems to support the hypothesis that reprise questions concern the semantic content of the fragment being reprised.

We have seen how an extension of G&C's contextual abstraction approach allows a corresponding analysis which holds to the strong version of this reprise content hypothesis, but also seen that standard HPSG analyses are not entirely consistent with the view of CNs as denoting predicates, and therefore would allow only the weaker version of the hypothesis to hold.

Examination of bare singular and plural CNs shows that mass nouns can be represented in a similar way (as denoting properties or kinds), but that some bare plurals must be represented differently, as individual referent reprise questions are possible.

The next section examines the implications of the content of reprise questions for the semantics of QNPs.

4.4 Noun Phrases

If we hold to the quantificational view of NP semantics, we should find that reprise questions concern a family of properties/sets (those properties which hold of the referent of the QNP). A referential view might instead lead us to expect that reprises of referential definites & specific indefinites should concern the individual referents directly.

4.4.1 Definite NPs

Taking a referential semantic viewpoint, we might therefore expect reprises of definite NPs to be paraphrasable as follows:

Clausal reading:

"Is it the individual X about which you are asking/asserting . . . X . . . ?"

Constituent reading:

"Which individual X do you intend to be referred to by the phrase NP?"

From a quantificational viewpoint, a paraphrase concerning a set of properties or sets might perhaps be expected:

Clausal reading:

"Is it the set of properties that hold of X about which you are asking/asserting . . . X . . . ?"

Constituent reading:

"Which set of properties do you intend to be conveyed by the phrase NP?"

Our corpus investigation included many types of definite NP: PNs, pronouns and demonstratives as well as definite descriptions. PNs have already been discussed in section 2.3.5,

and all examples found seemed perfectly consistent with that approach – we examine the others here. An overview of results is shown in table 4.3.

Demonstratives and Pronouns

Perhaps unsurprisingly (many of those who hold to the quantificational view believe demonstratives to be directly referential), our corpus investigation shows that demonstratives license the referential readings, not only when echoed verbatim as in example (144) (we shall call this kind of verbatim repeat a *direct echo*), but also when reprised with a co-referring PN as in examples (145) and (146), or with a reprise sluice as in examples (147) and (148). Both clausal and constituent versions seem available.

$(144)^{15}$

John:	Which way's North, do you know?
Sara:	That way.
John:	**That way?**
	Okay.

$\leadsto$ *"Are you telling me **that way there** is North?"*
$\leadsto$ *"By 'that way' do you mean that way there?"*

$(145)^{16}$

Christopher:	What was that lady <pause> <unclear>?
Dorothy:	**Julie?**
Christopher:	Mm.
Dorothy:	She's been with you, hasn't she?

$\leadsto$ *"Are you asking what **Julie** was <whatever>?"*
$\leadsto$ *"By 'that lady' do you mean Julie?"*

$(146)^{17}$

Brenda:	So have you seen this chap any more?
Jean:	**Mark?**
Brenda:	This <pause> new man <pause> is his name Mark?

$\leadsto$ *"Are you asking whether I've seen **Mark** any more?"*
$\leadsto$ *"By 'this chap' do you mean Mark?"*

---

[15]BNC file JP4, sentences 755–758
[16]BNC file KBW, sentences 883–886
[17]BNC file KBF, sentences 1228–1230

	Pattern	Referent	Functional	CN Predicate
Definite	*"... the N ..." / "The N?"*	10	6	2
		56%	33%	11%
Indefinite	*"... a(n) N ..." / "A(n) N?"*	-	-	28
		-	-	100%

Table 4.3: Literal Reprises – NPs

(147)[18]

Anon 1:	Oh God I hate these lot, they're so boring.
Cassie:	**What lot?**
Anon 1:	Them!
Cassie:	**Who?**
	What them lot?

⤳   *"What lot are you telling me you hate?"*
⤳   *"What lot do you mean by 'these lot'?"*

(148)[19]

Anon 1:	You'll have to speak to that boy again today.
Caroline:	**What boy?**
Anon 1:	Simon what's his name Steven Steve.
Caroline:	Oh right.

⤳   *"What boy are you telling me I'll have to speak to?"*
⤳   *"What boy do you mean by 'that boy'?"*

The same also appears to hold for pronouns, although we discuss these in more detail in section 4.5.2 below:

(149)[20]

Joanne:	It's, how many times did he spew up the stairs?
Emma:	**Julian?**
	Couple of times.

⤳   *"Is it **Julian**$_i$ that you are asking how many times i spewed up the stairs?"*
⤳   *"By 'he' do you mean Julian?"*

However, when we look at definite descriptions, the situation appears more complex: while referential readings are common, others are possible which do not appear to be directly referential.

---

[18]BNC file KP4, sentences 1546–1550
[19]BNC file KP3, sentences 2040–2043
[20]BNC file KCE, sentences 4190–4192

Definite Descriptions – Referential Readings

With definite descriptions, over half of the examples of direct echo questions found seemed to query the individual(s) being referred to.[21] Examples include constituent readings as in example (150) and clausal readings as in example (151):

(150)[22]	George:	You want to tell them, bring the tourist around show them the spot
	Sam:	**The spot?**
	George:	where you spilled your blood
	$\rightsquigarrow$	*"Which spot are you referring to by 'the spot'?"*

(151)[23]	John:	they'll be working on the, they'll be working on the kidnapper's instructions though wouldn't they?
		They would be working on the kidnapper's instructions, the police?
	Sid:	**The police?**
	John:	Aye
	Sid:	On
	Unknowns:	`<unclear>`
	Sid:	aye the, the senior detectives
	$\rightsquigarrow$	*"Is it **the police** who you are saying would be working . . . ?"*
	($\rightsquigarrow$	*"Who do you mean by 'the police'?"*)

Reprises using PNs As with demonstratives, definite descriptions can be reprised with another NP that conveys the same desired referent:

(152)[24]	Unknown:	And er they X-rayed me, and took a urine sample, took a blood sample. Er, the doctor
	Unknown:	**Chorlton?**
	Unknown:	Chorlton, mhm, he examined me, erm, he, he said now they were on about a slide `<unclear>` on my heart. Mhm, he couldn't find it.
	$\rightsquigarrow$	*"By 'the doctor' do you mean Chorlton?"*

---

[21]Comparison of the data in tables 4.1 and 4.3 shows that the reading distributions for definites and CNs are significantly different: a $\chi^2_{(1)}$ test shows that the probability p that the referent/predicate reading distribution is *independent* of whether the source is a definite NP or a CN is tiny ($p < 0.01\%$). The difference between the distributions for definites and indefinites is similarly significant ($p < 0.01\%$). There is no significant difference between indefinites and CNs, however, as discussed in section 4.4.2.

[22]BNC file KDU, sentences 728–730

[23]BNC file KCS, sentences 660–665

[24]BNC file KPY, sentences 1005–1008

(153)[25]

Brian:	Have you seen the advert?
John:	Which one?
Brian:	With the bull dog `<pause>` tha– , you know the dog you like
John:	**George?**
Brian:	Yeah.
John:	Yeah.
Brian:	Well `<pause>` the husband goes out `<pause>` and she's got his dinner and he's underneath sort of going

⤳   *"By 'the dog you like' do you mean George?"*

This is interesting: not only does it give further weight to the idea that these reprises are genuinely referential (PNs are generally held to be referential even by those who hold to the quantificational view of definite NPs), it also suggests that the referent can be an entity in the world (rather than some kind of discourse object).

Sluices   And again, reprise sluices are available which seem to concern a referent:

(154)[26]

Terry:	Richard hit the ball on the car.
Nick:	**What car?**
Terry:	The car that was going past.
Nick:	**What ball?**
Terry:	James [`last name`]'s football.

⤳   *"Which car are you saying Richard hit the ball on?"*
⤳   *"Which car do you mean by 'the car'?"*

⤳   *"Which ball are you saying Richard hit on the car?"*
⤳   *"Which ball do you mean by 'the ball'?"*

(155)[27]

Heidi:	How's the box going?
Vicki:	**Which box?**
Heidi:	The new one.
Vicki:	Oh that one. `<pause>`

⤳   *"Which box$_x$ are you asking how x is going?"*
⤳   *"Which box do you mean by 'the box'?"*

Referential Analysis   Two points are perhaps worth reinforcing: firstly, definite descriptions, pronouns, demonstratives and proper names all seem to make the same kind of referen-

---

[25]BNC file KCL, sentences 3833–3839
[26]BNC file KR2, sentences 862–866
[27]BNC file KC3, sentences 1377–1380

tial reprise questions available; secondly, it seems very hard to interpret any of these examples as querying a family of sets (a GQ) rather than an individual referent.

It also seems difficult to reconcile these examples with the Kripkean view of reference via pragmatics as outlined in section 4.2. Examples like example (152), in which a referential question is asked (and answered) before the sentence containing the original NP has been finished (indeed, it has hardly been started) do not obviously permit an explanation which requires understanding of the proposition expressed as an early step.[28] Secondly, if what is being reprised is the result of pragmatic inference from a GQ, why do readings querying the GQ itself and other associated inferences not seem to be available?

A proposal with more explanatory power therefore seems to be that the content of definite NPs must at least contain, and perhaps consist entirely of (as sketched out roughly in AVM (156) – we will fill in the details in section 4.4.4), the intended referent (which in the case of plurals, we assume will be a set). An analysis of these referent reprise questions would then be available along identical lines to G&C's analysis for PNs – an identifiable referent for the contextual parameter must be found in context as part of the grounding process.[29]

$$(156) \quad \begin{bmatrix} \text{PHON} & \langle \text{the, dog} \rangle \\ \text{CONTENT} & \boxed{1}\big[x : the\_dog(x) \big] \\ \text{C-PARAMS} & \big\{ \boxed{1} \big\} \end{bmatrix} \qquad \lambda\big\{ x \big\}[the\_dog(x)].x$$

Definite Descriptions – Functional Readings

Most of the rest of the examples of direct echoes of definite descriptions did not seem to be querying an individual referent, but rather seemed to be querying a function or its domain. As might be expected, these examples were mostly *attributive* uses, which have long been held up as examples against the referential nature of definite descriptions (e.g. Russell's examples *"the centre of mass of the solar system"*, *"the first person born in the twenty-first century"*[30]), but other types that we would expect to behave in this way include *de dicto* uses, *narrow scope* uses, Poesio (1994)'s *weak* definites, and *generic* uses, none of which obviously convey direct reference.

Following Barwise and Perry (1983), the function expressed by attributive uses can be taken to be one from situations to individuals. Example (157) (taken from an oral history interview and describing typical activities on a Scottish estate) shows a question which seems not to query the identity of the actual referents (the individual pools at a particular time) but rather the identity of the function and thus the distinguishing feature of the pools which

---

[28]Minimally, it would require a radically incremental view of semantic processing.

[29]Whether this process should be restricted to allow grounding only to unique referents, or to most familiar or most salient referents, will not be addressed here, but is not essential to the basic approach.

[30]Of course, Russell was writing nearly a hundred years ago. We could conceivably use this example referentially now.

required maintenance at any particular time.

(157)[31]

	Anon 1:	In those days how many people were actually involved on the estate?
	Tommy:	Well there was a lot of people involved on the estate because they had to repair paths.
		They had to keep the river streams all flowing and if there was any deluge of rain and stones they would have to keep all the pools in good order and they would
	Anon 1:	**The pools?**
	Tommy:	Yes the pools. That's the salmon pools
	Anon 1:	Mm.
	$\leadsto$	*"What are you intending 'the pools' to pick out in the situation you are describing?"*
	$\leadsto$	#*"Which actual entities are you referring to by 'the pools'?"*

Similarly in example (158) (taken from a music lesson), the question (from the student) does not seem to be asking about the actual musical note names required (as that is what (s)he is being asked to produce as an answer), but rather seems to have an argument or domain reading available (about the particular notes or music for which the names are required – amongst other possibilities):

(158)[32]

	Eddie:	I'm used to sa–, I'm used to being told that at school.
		I want you \<pause\> to write the names of these notes up here.
	Anon 1:	**The names?**
	Eddie:	The names of them.
	Anon 1:	Right.
	$\leadsto$	*"What situation/notes are you intending me to interpret 'the names' relative to?"*
	$\leadsto$	?*"What are you intending 'the names' to refer to in that situation?"*
	$\leadsto$	#*"Which actual names are you referring to by 'the names'?"*

Again, a reading concerning properties of properties or sets of sets does not seem plausible. A reasonable proposal which captures such uses might therefore be an analysis as sketched in AVM (159), this being the functional equivalent of the version in AVM (156) above, with its constituent function and argument becoming the abstracted parameters in

---

[31]BNC file K7D, sentences 307–313

[32]BNC file KPB, sentences 417–421

C-PARAMS:

(159)
$$
\begin{bmatrix}
\text{PHON} & \left\langle \text{the, dog} \right\rangle \\
\text{CONTENT} & \left[f(s) : \boxed{1}[f : f = \mathit{the\_dog}], \boxed{2}[s : s \subseteq DOM(f)] \right] \\
\text{C-PARAMS} & \left\{ \boxed{1}, \boxed{2} \right\}
\end{bmatrix}
$$

(160) $\lambda\{f, s\}[f = \mathit{the\_dog}, s \subseteq DOM(f)].f(s)$

Grounding therefore requires both the function f and the argument d to be found in context. Failure to do so would therefore license clarification questions which can be read as concerning either function or argument/domain, or both. Note that the job of identifying the domain corresponds to Poesio (1993)'s view of definite interpretation as anchoring a parameter corresponding to the resource situation, but that on the view presented here this is not *all* that is required.

The domain of the function need not necessarily be one of situations: indeed, for narrow-scope definites it seems simpler to take the domain as being a set of individuals contributed by a wider-scoping NP (and this is set out in section 4.5.1). However, the treatment of the semantic content as functional, with the resulting contribution to C-PARAMS, remains.

Strong/Weak Hypothesis   This representation does not fit exactly with the strong version of the reprise content hypothesis as it is currently phrased. While both constituent elements of the content (function and argument) are reprisable, a single question might of course query only one of them, thus holding only to the weak version of the hypothesis. However, querying the entire content directly would seem wrong here, as it would necessarily reduce the functional representation to the non-functional version.

This raises the possibility that a third, intermediate stance is what is actually required. One intuitively attractive version might be that reprise questions can query *any* part of the reprised fragment's semantic content (including the content as a whole). This is weaker than the strong version, as questions are no longer restricted to only ask about the whole content; and it is stronger than the weak version as it would require all parts of the content (including the whole) to be queriable, rather than just some part. It would also fit with a functional representation, as it would allow either function or argument to be reprised. However, it is immediately problematic, as it is not clear what constitutes a "part": function and argument seem reasonable, but what about smaller constituent sub-formulae? As we will see in the next section, reprises do seem to be able to focus on syntactic sub-constituents – but it is not at all clear to how small a scale this might extend. For now, then, it seems safest to maintain the original strategy of proposing representations which stick to the strong hypothesis whenever possible, but recognising that there are some cases (such as functional NPs) where some weakening will be required.

Ambiguity   Introduction of this alternative analysis means, of course, assuming some ambiguity in the representation of definites: but note that this is not an ambiguity of semantic type (the content is still of type e). This ambiguity could be removed by taking *all* definite descriptions to be functional, with referential definites those where the situational argument s is the current utterance situation s_0 (thus resembling von Heusinger (2002)'s analysis of specific indefinites as those functional on the speaker).[33] In such cases, grounding of the function f in the known current situation s_0 is equivalent to identifying the referent $x = f(s_0)$. As this appears to be a worst-case analysis (over half of the corpus examples appeared to be directly referential), this step is not taken here, but merely noted as an option.

It seems likely that such a step would not be required for PNs and demonstratives in any case, which do not appear to have functional versions (not being able to take narrow scope),[34] so these would keep the previous simple referential analysis.

Definite Descriptions – Sub-Constituent Readings

The few remaining examples of definite NP reprises found seemed to be easier to interpret as having a predicate reading, identical to that which would be obtained by reprising the CN alone. No intonational information is available in the BNC, but these readings appear to be those that are made more prominent by stressing the CN (see example (161)).

(161)[35]	Anon 1:	They'd carry the sack on their back?
	George:	On the back, the bushel, yes
	Anon 1:	**The bushel?**
	George:	`<unclear>`
	Anon 1:	`<unclear>`
	George:	The corn.

$\leadsto$ *"What are you referring to by 'the bushel'?"*
$\leadsto$ *"What property do you mean by 'bushel'?"*
$\leadsto$ *"Is it the thing with the property **bushel** that you're saying . . . "*

This does not seem to be restricted to definites: in fact, the same readings appeared to be possible for all other NPs we examined (as we will see below). It therefore seems reasonable to assume that this reading is in fact a focussed reprise of the CN rather than a question about the NP as a whole. Examination of sluices reinforces this: where reprise sluices were found

---

[33]Of course, removing this ambiguity here would lead to more work later. When resolving scope, there will be more arguments which need their reference established – see section 4.5.1.

[34]Although possible counterexamples have been proposed for demonstratives – see (Roberts, 2002).

[35]BNC file H5H, sentences 254–257

with this reading, only the CN was substituted by a *wh*-word, rather than the whole NP:

(162)[36]

Elaine:	what frightened you?
Unknown:	The bird in my bed.
Elaine:	**The what?**
Audrey:	**The birdie?**
Unknown:	The bird in the window.

$\leadsto$ *"What property$_x$ is it you're saying the thing with x frightened you?"*

Similarly, although none were found in the BNC, it seems plausible that a reading corresponding to the logical relation expressed by the determiner is possible (again, the reader may find this easier to capture by imagining intonational stress on the determiner).

In other words, the readings available for reprises of sub-constituents of the NP are still available when reprising the NP, especially when the relevant sub-constituent is stressed. This might be expected, given the idea of C-PARAMS inheritance outlined in section 2.3.5. The reprise content hypothesis must therefore be re-formulated to allow for these "inherited" daughter questions:

Reprise Content Hypothesis (revised weak version):

(163)   *A reprise fragment question queries part of the standard semantic content of the fragment being reprised or one of its syntactic daughters.*

Reprise Content Hypothesis (revised strong version):

(164)   *A reprise fragment question queries exactly the standard semantic content of the fragment being reprised or one of its syntactic daughters.*

This has implications for exactly how C-PARAMS inheritance should be reflected in the grammar, and also requires a theory of sub-constituent focussing to explain how the readings arise (see section 4.5.4).

4.4.2   Indefinite NPs

So we have seen that the evidence provided by reprises of definite NPs leads us towards a view of them as referential (although possibly functional) rather than quantificational. In this section, we turn to *indefinites*. Again, a referential viewpoint might lead us to expect that reprises of indefinites should involve a referent (perhaps not a specific real-world object but a

---

[36]BNC file KBC, sentences 1193–1197

discourse referent (Kamp and Reyle, 1993), belief object (Zimmerman, 1999) or intentional object (Dekker, 2002)), and that this referent would therefore be queried by a reprise question.

Sub-Constituent Readings

However, if they do exist, such readings seem to be uncommon. All direct echo examples found were most felicitous when read as the sub-constituent readings described in section 4.4.1 above. For plain singular indefinites (see table 4.3), all examples seemed identical to the CN predicate reading (whether clausal or constituent). Note that the constituent reading, paraphrased in the examples below as *"What property do you mean by 'N'?"*, might also be paraphrased *"What is a N?"* – but that this should not be confused with a constituent reading which asks about the whole NP reference, *"Which N do you mean by 'a N'?"*.

$(165)^{37}$

Mum:	What it ever since last August. I've been treating it as a wart.
Vicky:	**A wart?**
Mum:	A corn and I've been putting corn plasters on it
$\rightsquigarrow$	*"Is it the property **wart**$_i$ that you're saying you've been treating it as something with i?"*
$\rightsquigarrow$	*"What property do you mean by 'wart'?"*
$\rightsquigarrow$	#*"Which wart are you saying you've been treating it as?"*

$(166)^{38}$

Unknown:	What are you making?
Anon 1:	Erm, it's a do– it's a log.
Unknown:	**A log?**
Anon 1:	Yeah a book, log book.
$\rightsquigarrow$	*"Is it the property **log** that you're saying it's something with?"*
$\rightsquigarrow$	*"What property do you mean by 'log'?"*
$\rightsquigarrow$	#*"Which log are you saying it is?"*

For plural indefinites the same holds, although a reading querying the determiner rather than the predicate is also available (as was suggested might be possible for definites in sec-

---

[37]BNC file KE3, sentences 4678–4681
[38]BNC file KNV, sentences 188–191

tion 4.4.1 above):

$(167)^{39}$

Lara:	There's only two people in the class.
Matthew:	**Two people?**
Unknown:	For cookery, yeah.
Lara:	Yeah `<unclear>` `<laugh>`.

$\rightsquigarrow$ *"Is it **two**$_N$ that you're saying there's N people?"*
$\rightsquigarrow$ *"Is it **people** that you're saying there's two of?"*
$\rightsquigarrow$ #*"Which two people are you saying are in the class?"*

$(168)^{40}$

Anon 2:	Was it nice there?
Anon 1:	Oh yes, lovely.
Anon 2:	Mm.
Anon 1:	It had twenty rooms in it.
Anon 2:	**Twenty rooms?**
Anon 1:	Yes.

$\rightsquigarrow$ *"Is it **twenty**$_N$ that you're saying it had N rooms?"*
$\rightsquigarrow$ *"Is it **rooms** that you're saying it had twenty of?"*
$\rightsquigarrow$ #*"Which twenty rooms are you saying are it had?"*

Two approaches therefore present themselves: either the content of an indefinite (be it referential or quantificational) is simply not abstracted to the C-PARAMS set, thus leaving only parameters associated with sub-constituents to be reprised; or the content of an indefinite is in fact identical to that of one of its sub-constituents. The second seems problematic: firstly, which sub-constituent would we choose? As seen above (e.g. in example (168)), both determiner and CN content seem to be available. Secondly, it would mean different semantic types for definites and indefinites. There are other problems too, not least for an account of anaphora (see section 4.5.2 below for more details). In any case, the argument for making this step does not seem strong: after all, the same sub-constituent questions are available for definites.

Sluices This is perhaps reinforced by the fact that reprise sluices which query the CN predicate seem to be equally common for definites and indefinites. As shown in table 4.4, the same number of *"A what?"* reprises (see example (169) below) were found as *"The what?"* reprises (see example (162) above).[41] This is hardly strong evidence, but might help us to believe that sub-constituent questions are no *more* made available by indefinites than definites, as one might expect them to be if the content of indefinites really was the same as that of one

---

[39] BNC file KPP, sentences 352–355

[40] BNC file K6U, sentences 1493–1498

[41] Although definites are more common than indefinites in the BNC (nearly twice as many), there is no statistically significant difference between the relative numbers of predicate sluices shown in table 4.4 and the relative numbers of overall occurrences.

of their sub-constituents.

(169)[42]

Stuart:	I know it's good in it? `<unclear>` but erm, `<unclear>` bought her, I've bought her a Ghost video.
Mark:	**A what**?
Stuart:	A Ghost video.
Mark:	Oh yeah.

⤳   *"What property P is it you're saying you've bought her something with P?"*
⤳   *"What property do you mean by 'Ghost video'?"*
⤳   #*"Which Ghost video are you saying you've bought her?"*

	Pattern	Number in BNC
Definite	*"… the N … " / "The what?"*	10
Indefinite	*"… a(n) N … " / "A(n) what?"*	10

Table 4.4: Predicate Sluices

It therefore seems more reasonable to take the first approach: that indefinite content is not easily available for reprise, and so sub-constituent readings predominate. But in that case, can we shed any light on whether a referential or quantificational analysis better explains the facts?

Possible Referential Readings

While no clear examples were found in our corpus study, we feel that there *is* a possibility of referential questions with specific indefinites where the hearer realises that the speaker has a particular referent in mind, and intends the hearer to be able to identify it (what Ludlow and Segal (2004) call *definite* indefinites). Some BNC examples, while most felicitous when read as CN predicate queries, do seem to offer a possible referential paraphrase:

(170)[43]

Stefan:	Everything work which is contemporary it is decided
Katherine:	Is one man?
Stefan:	No it is a woman
Katherine:	**A woman?**
Stefan:	A director who'll decide.
Katherine:	She's good?
Stefan:	Hm hm very good.

⤳   *"Is it a **woman** you are saying it is?"*
⤳   ?*"Which woman are you saying it is?"*

---

[42]BNC file KDA, sentences 672–675
[43]BNC file KCV, sentences 3012–3018

$(171)^{44}$

	Skonev:	there's a heron in our garden
	Patrick:	**a heron?** where?
	Skonev:	just round the side
	Patrick:	Is it dead?
	$\rightsquigarrow$	*"Is it a **heron** you are saying is in our garden?"*
	$\rightsquigarrow$	*?"Which heron? where?"*

Sluices   If this is the case, we should expect referential reprise sluices *"What/Which N?"* (as opposed to the CN predicate sluice *"A what?"* described above) to be available, if rare. We already know that this kind of reprise sluice, if it exists, must be rare: section 3.3 did not identify any indefinite NPs as sources of reprise sluices in the sub-corpus of CRs, in contrast to definites and CNs. Searching through the whole BNC for *"Which N?"* examples shows that they certainly exist for indefinites, and are indeed rare (about 6 times less common after *a N* than after *the N* – see table 4.5).^{45}

	Pattern	Number in BNC
Definite	*"... the N ... " / "What/Which N?"*	25
Indefinite	*"... a(n) N ... " / "What/Which N?"*	4

Table 4.5: Referential Sluices

However, we must be careful when examining these examples, as it is important to distinguish between *reprise* sluices – questions concerning the directly conveyed content of the utterance, asked by the hearer during the comprehension (grounding) process, and typically delivered with a rising reprise intonation – and the more familiar *direct* sluices – questions asking for more specific information than that directly conveyed, which are not asked during the comprehension process but can be asked even after complete acceptance of an assertion, and which do not appear with the same rising reprise intonation. The former ask about part of the content which was intended to be conveyed in the source utterance, the latter do not.

Of course, especially given the lack of intonational information in the BNC, it is very difficult to determine the reprise/direct nature of a sluice beyond any doubt – we can merely attempt to fit plausible paraphrases to the dialogue context. In most cases (see examples (172) and (173)), both interpretations seemed plausible, although the direct version arguably more likely. But one example in particular (example (174)) seemed to support a reprise reading more readily: the speaker appears to be using an indefinite in order to identify a person

---

^{44}BNC file KR1, sentences 230–233

^{45}The referential sluice distribution between definites and indefinites (table 4.5) is significantly different from the predicate sluice distribution (table 4.4): a $\chi^2_{(1)}$ test shows probability of independence $p < 1\%$. It is also not merely an effect of the fact that definites are more common ($p < 2\%$).

without mentioning him by name, while the interviewer wants to be sure he has understood the intended reference correctly.

(172)[46]

Nicola:	We're just going to Beckenham because we have to go to a shop there.
Oliver:	**What shop?**
Nicola:	A clothes shop. `<pause>` and we need to go to the bank too.

⤳   reprise: *"(Sorry,) What shop are you telling me we have to go to?"*
⤳   reprise: *"(Sorry,) What shop do you mean 'a shop' to refer to?"*
⤳   direct: *"(OK, I see.) What shop is it we have to go to?"*

(173)[47]

Damion:	Give me it.
Terry:	Why?
Damion:	I want you to show it someone.
Terry:	**Who?**
Damion:	A girl in my form.
Terry:	**Who?**
	Why?
Damion:	I'll give it to you later.

⤳   reprise: *"(Sorry,) Who are you telling me you want me to show it to?"*
⤳   reprise: *"(Sorry,) Who do you mean by 'someone'?"*
⤳   direct: *"(I see.) Who do you want me to show it to?"*

(174)[48]

Ray:	And of course, when this all happened, and I'm listening to what people are saying tonight, it's it's sort of making me feel a bit sick what they're saying.
Nicky Campbell:	Why is that?
Ray:	One supports that I lay in the street looking and waiting for a a man they mention tonight and that man is a well known killer of British soldiers. And I'm now asked
Nicky Campbell:	**Which man?**
Ray:	I'm now asked to respect him. And I'm sorry, I cannot respect a man
Nicky Campbell:	**The man who's name has been mentioned tonight?**
Ray:	Tonight. I cannot say that anybody can respect a man in this country and to run for their country as a well known I R A supporter. And he's up there on one of your pictures.
Nicky Campbell:	Mhm.

⤳   reprise: *"(Sorry,) Which man do you mean by 'a man they mention tonight'?"*
⤳   reprise: *"(Sorry,) Which man are you telling me you lay waiting for?"*
⤳   direct: *"(I see.) Which man did you lie waiting for?"*

---

[46]BNC file KDE, sentences 2214–2217
[47]BNC file KR2, sentences 1533–1540
[48]BNC file HV2, sentences 225–236

Again, no examples seemed to support a property-of-properties or set-of-sets paraphrase at all. Taken together with the possible referential fragment reprises (examples (170) and (171)), this is at least tentative support for a view that indefinites (a) are better seen as referential than quantificational, and (b) that this referential term can in certain cases be contextually abstracted, thus being available for reprise questions. An analysis of indefinites should therefore allow for such readings to be constructed: as for definites, their content should consist (at least in part, and if holding to the strong hypothesis, entirely) of an individual or set of individuals. The distinction from definites is that in ordinary uses this content is not contextually abstracted, and therefore does not have to be identified during grounding (leaving only the sub-constituents which must be grounded and can be clarified), but instead must be existentially quantified within the sentence (see AVM (175) for a sketch – existential quantification is achieved by membership of the STORE feature; more details are given in section 4.5.1). Definite uses are distinguished simply by making the content a member of C-PARAMS as in AVM (176), so that it does have to be grounded in context, and can be reprised.

$$(175) \quad \begin{bmatrix} \text{PHON} & \left\langle \text{a, dog} \right\rangle \\ \text{CONTENT} & \boxed{1}\left[x:dog(x)\right] \\ \text{STORE} & \left\{ \boxed{1} \right\} \\ \text{C-PARAMS} & \{\} \end{bmatrix} \qquad \exists\left\{x\right\}[dog(x)].x$$

$$(176) \quad \begin{bmatrix} \text{PHON} & \left\langle \text{a, dog} \right\rangle \\ \text{CONTENT} & \boxed{1}\left[x:dog(x)\right] \\ \text{STORE} & \{\} \\ \text{C-PARAMS} & \left\{ \boxed{1} \right\} \end{bmatrix} \qquad \lambda\left\{x\right\}[dog(x)].x$$

This view of indefinites as individuals which are existentially quantified (rather than as generalized quantifiers) is not dissimilar to the *choice function* approach of Reinhart (1997); Szabolcsi (1997), or the *epsilon term* approach of van Rooy (2000); von Heusinger (2000); Kempson et al. (2001) – where indefinites denote individuals chosen by some existentially quantified choice function. While these approaches seem perfectly consistent with the observations here, for simplicity's sake the representations proposed here will quantify over the individuals directly, although functional versions will be used to express relative scope in section 4.5.1 below.

This account also allows an analysis of sluicing which expresses the distinction between direct and reprise sluices: direct sluices are those which concern an existentially quantified referent contributed by a previous grounded utterance (essentially the analysis of SHARDS and G&S); while reprise sluices are those which concern the identity of a member of C-PARAMS during grounding, following G&C.

4.4.3 Other Quantified NPs

We have so far only considered definite and indefinite NPs. What of QNPs which contain
other quantifiers?

There are really very few examples of reprises of such QNPs in the BNC[49], so it is pre-
mature to claim strong results; but what indications we could get, together with our intuition,
point towards an identical analysis to that proposed above for indefinites. Most examples
seem most felicitous when interpreted as concerning sub-constituents (either the CN predi-
cate or the logical relation expressed by the quantifier), but seem to have a possible referential
interpretation too:

(177)[50]

	Richard:	No I'll commute every day
	Anon 6:	**Every day?**
	Richard:	as if, er Saturday and Sunday
	Anon 6:	And all holidays?
	Richard:	Yeah `<pause>`
	$\rightsquigarrow$	*"Is it **days**$_N$ that you are saying you'll commute every N?"*
	$\rightsquigarrow$	*"Is it **every** day that you are saying you'll commute?"*
	$\rightsquigarrow$	*"Which days do you really mean by 'every day'?"*

With universals as in example (177) above, we should perhaps not be surprised by refer-
ential readings: it has been suggested that universals should be considered as definites (see
e.g. Prince, 1992; Abbott, 2003). They are less clearly available with other quantifiers:

(178)[51]

	Anon 1:	Er are you on any sort of medication at all Suzanne?
		Nothing?
	Suzanne:	No. Nothing at all.
	Anon 1:	**Nothing?**
		No er things from the chemists and cough mixtures or anything `<unclear>`?
	$\rightsquigarrow$	*"Is it **no** things that you are saying you're on?"*
	$\rightsquigarrow$	?*"Which things do you really mean by 'nothing'?"*

As before, examples seem to be possible where referential uses can be made more clear
by use of co-referring PNs in the reprise, although this time we have to rely on imagined

---

[49]This is not surprising, as these NPs are relatively rare in the BNC to begin with. They are an order of
magnitude less common than *"the/a N"*: there are more than 50 times more sentences containing *"the N"* as there
are containing *"every N"*, and *"most N"*, *"many N"* and *"few N"* are even rarer. As we found fewer than 100
reprises of *"the N"*, we would only expect a few *"every N"* reprises, and none for the other quantifiers, and this
is what we find.

[50]BNC file KSV, sentences 257–261

[51]BNC file H4T, sentences 43–48

examples:

(179)

A:	I want everyone in here to come with me.
B:	**Everyone? / Me, Carl and Donna?**

$\leadsto$ *"Who do you mean by 'everyone'?"*
$\leadsto$ *"By 'everyone' do you mean B, C and D?"*

(180)

A:	Most people came to the party.
B:	**Most people?**
A:	Well, me, Brenda and Carmen.

$\leadsto$ *"Who do you mean by 'most people'?"*

This possibility suggests that as for indefinites, these QNPs should be analysed as existentially quantified sets of individuals, which are not contributed to C-PARAMS under normal circumstances. Referential uses are obtained simply by adding the content to C-PARAMS. In the next section, we outline this approach in more detail.

4.4.4   Semantic Analysis

If we are to hold to the reprise content hypothesis, the availability of referent readings for QNP reprise questions means that the semantics of QNPs must (at least partially) consist of a referent individual or set. It seems clear that this referent is the witness set of the corresponding GQ (where this set may be functionally dependent on a situation or another set).

Two approaches present themselves. Firstly, we can hold to a standard view of QNPs as denoting GQs, and assume that the witness set forms the parameter to be grounded in context. This will, of course, only hold to the weaker version of our hypothesis. Secondly, as we have been sketching out so far, we can hold to the stronger version by considering QNPs to denote witness sets directly.

QNPs as GQs

The first approach is shown in (182) for the definite NP *the dog*. The content is a GQ, and the abstracted parameters which must be grounded are the witness set w (containing the referent dog to be identified in context) and the parameters contributed by the sub-constituents – the predicate P denoted by the CN *dog* and the logical relation Q denoted by the determiner.[52]

---

[52]Requiring determiner relations to be grounded in context may seem counterintuitive, but will be discussed in more detail in section 4.6.1 – for now, perhaps it is enough to note that determiners can certainly be reprised, and that the questions asked seem able to concern a logical relation which is surprising in context:

(181)

A:	Most/two/only a few students came to the party.
B:	**Most/Two/Only a few?**

An equivalent indefinite version would of course not add the witness set to the abstracted C-PARAMS set, leaving only the sub-constituent parameters.

(182) $\lambda \{w, Q, P\} [witness(w, Q(P)), Q = the, name(P, dog)].Q(P)$

The relation $witness(w, Q(P))$ is of course defined as:

(183) $witness(w, Q(P)) \quad \leftrightarrow \quad w \subseteq P \wedge w \in Q(P)$

This would account for the availability of referential reprise questions: failure to find a suitable witness set in context will result in a clarification question concerning its identity. This solution, however, only holds to the weak version of the reprise content hypothesis, as the reprise question would no longer concern the entire content of the NP, but only a part. As such, it does not offer a clear explanation of why reprise questions can only query this part, rather than the whole GQ content.

QNPs as Witness Sets

Accordingly the second approach seems preferable: to treat QNPs as denoting their witness sets directly. This leads to a simple representation, using B&C's equivalence stated in section 4.2.3 above, that a verbal predicate holds of a QNP iff the witness set belongs to the set expressed by that predicate.[53] The content is therefore a set, which for definites is also a member of the set of contextually abstracted parameters, along with those contributed by sub-constituents:

(184) $\lambda \{w, Q'', P\} [w = Q''(P), Q'' = the'', name(P, dog)].w$

Here the function the'' which picks out our witness set is defined via the following equivalences:

(185) $w = Q''(P) \quad \leftrightarrow \quad Q'(w, P) \quad \leftrightarrow \quad witness(w, Q(P))$

Essentially this will give a semantic representation of a sentence *"the dog snores"* which can be written as follows:

(186) $the'(w, P) \wedge dog(P) \wedge snore(w)$

which is broadly similar to the representation of (Hobbs, 1983, 1996).[54] Following B&C's equivalence, the sentence is true iff $w \subseteq snore$.

This solution has the same power to account for clarifications as the previous one (the witness set forms the contextual parameter to be grounded), but also holds to the strong version of our reprise content hypothesis, and therefore straightforwardly explains why reprise ques-

---

[53]This could alternatively be thought of as implicitly universally quantifying over the members of the witness set.

[54]Although Hobbs uses the notion of a *typical element* of a set and uses this as the argument of a verb (coercing the predicate into a typical/non-typical version as necessary). This step is not taken here.

tions can only concern this set (or a sub-constituent). However, this version holds only for MON$\uparrow$ quantifiers: some possible solutions for other quantifiers are discussed in section 4.5.3 below.

For functional versions, the representation is similar, but function and argument must be separately abstracted, and defined as giving a witness set:

(187) $\lambda\{f, s, Q''', P\}[f = Q'''(P), s \subseteq DOM(f), Q''' = the''', name(P, dog)].f(s)$

where

(188) $f = Q'''(P) \quad \leftrightarrow \quad \forall x.x \in DOM(f) \rightarrow witness(f(x), Q(P))$

4.4.5 HPSG Analysis

We are now in a position to give a HPSG analysis which shows how the NP's semantic representation is built up from those of its daughters. However, it turns out to be slightly at odds with the usual head-driven principles of HPSG: neither CONTENT nor C-PARAMS is now being directly inherited from or amalgamated across syntactic daughters.[55]

CONTENT **Specification**

As pointed out in section 4.3.3 above, holding to the strong version of our reprise content hypothesis must mean that NPs do not inherit their content from their head daughter CNs (as in standard HPSG unification-based semantics), or simply amalgamate across daughters (as in Minimal Recursion Semantics): the referential reprises available for NPs are simply not available when reprising the daughters. To specify the content correctly, we must therefore

---

[55]While an analysis is given here only for the preferred witness-set only approach, the general observations also hold for the GQ approach.

posit a type *qnp* for all QNPs which specifies how the semantic representation is built:

$$
(189) \quad
\begin{bmatrix}
qnp \\
\text{CONTENT} \begin{bmatrix}
\text{INDEX} & w \\
\text{RESTR} & \left\{ \begin{bmatrix}
witness\_set\_rel \\
\text{INSTANCE} & w \\
\text{PROPERTY} & P \\
\text{RELN} & Q
\end{bmatrix} \right\}
\end{bmatrix} \\
\text{DTRS} \quad \left\langle \begin{bmatrix} det \\ \text{CONTENT} \,|\, \text{INDEX} & Q \end{bmatrix}, \begin{bmatrix} nominal \\ \text{CONTENT} \,|\, \text{INDEX} & P \end{bmatrix} \right\rangle
\end{bmatrix}
$$

(or in abbreviated form):

$$
(190) \quad
\begin{bmatrix}
\text{CONTENT} & \begin{bmatrix} w : w = Q(P) \end{bmatrix} \\
\text{DTRS} & \left\langle \begin{bmatrix} det \\ \text{CONTENT} & Q \end{bmatrix}, \begin{bmatrix} nominal \\ \text{CONTENT} & P \end{bmatrix} \right\rangle
\end{bmatrix}
$$

Note that the constraint expressed above is still monotonic (no semantic information is dropped in construction of the mother) and compositional (the semantics of the mother is obtained purely by functional application of daughters).

C-PARAMS **Amalgamation**

As mentioned in section 4.4.1 above, the availability of sub-constituent readings shows that the C-PARAMS value for a phrase must include the values of its daughters. However, the fact that reprises of head daughters (e.g. CNs) cannot be interpreted as querying the content of their *sisters* (e.g. determiners) means that this inheritance process cannot be via lexical heads (as in the general Non-LOCAL Amalgamation Constraint assumed to govern C-PARAMS by G&C), but instead must be explicitly specified for the mother. It could be expressed instead as a default constraint on the type *phrase* similar to G&C's CONSTITS Amalgamation Constraint, shown in AVM (191) below:

$$
(191) \quad
\begin{bmatrix}
phrase \\
\text{C-PARAMS} & \boxed{1} \cup \ldots \cup \boxed{n} \\
\text{DTRS} & \left\langle \begin{bmatrix} \text{C-PARAMS} & \boxed{1} \end{bmatrix}, \ldots, \begin{bmatrix} \text{C-PARAMS} & \boxed{n} \end{bmatrix} \right\rangle
\end{bmatrix}
$$

However, definite NPs (which we can now take to include the referential uses of indefinites and other QNPs) would have to override this default, as they introduce a new contextual parameter as well as amalgamating those of their daughters. Standard indefinites would hold to it, but we must ensure that their content is instead existentially quantified. These facts can be combined into a general *definiteness principle*.

Definiteness Principle

In HPSG terms, indefinites must contribute their content to the STORE feature (which specifies the existentially quantified elements – see section 4.5.1 for more details), while definites contribute it to C-PARAMS. It is precisely this that distinguishes definite from indefinite uses. We can therefore state a general principle: the content of a NP must be a member of either C-PARAMS or STORE. We can replace AVM (191) with a more general Definiteness Principle, which applies to both words and phrases. For words, it is simply expressed:

$$
(192) \quad
\begin{bmatrix}
word \\
\text{CONTENT} & \boxed{1} \\
\text{STORE} & \boxed{2} \\
\text{C-PARAMS} & \left\{\boxed{1}\right\} - \boxed{2}
\end{bmatrix}
$$

For phrases, it must also specify STORE and C-PARAMS inheritance from daughters. The C-PARAMS value of the mother is the union of the daughter values, plus the mother content, unless this is contributed to STORE. As shown here, this is currently restricted to noun phrases (phrases whose head is of type *nominal*), as we have no evidence that it applies to other types:

$$
(193) \quad
\begin{bmatrix}
phrase \\
\text{CAT} \mid \text{HEAD} & nominal \\
\text{CONTENT} & \boxed{1} \\
\text{STORE} & \boxed{2} \cup \boxed{3} \\
\text{C-PARAMS} & (\left\{\boxed{1}\right\} - \boxed{2}) \cup \boxed{4} \cup \ldots \cup \boxed{n} \\
\text{HEAD-DTR} & \begin{bmatrix} \text{STORE} & \boxed{3} \end{bmatrix} \\
\text{DTRS} & \left\langle \begin{bmatrix} \text{C-PARAMS} & \boxed{4} \end{bmatrix}, \ldots, \begin{bmatrix} \text{C-PARAMS} & \boxed{n} \end{bmatrix} \right\rangle
\end{bmatrix}
$$

Referential phrases (like definites) can therefore be specified as inheriting their STORE values directly from their head daughters, with referential words (such as CNs, which on this account are referential to a predicate) specified as having empty STORE, thus forcing the content of both to be a member of C-PARAMS – see AVM (194). Non-referential phrases (like indefinites) can be specified as contributing to STORE, and thus can make no contribution to C-PARAMS, as shown in AVM (195).

$$
(194) \quad
\begin{bmatrix}
phrase~\&~referential \\
\text{STORE} & \boxed{1} \\
\text{HEAD-DTR} \mid \text{STORE} & \boxed{1}
\end{bmatrix}
\qquad
\begin{bmatrix}
word~\&~referential \\
\text{STORE} & \{\}
\end{bmatrix}
$$

$$
(195) \quad
\begin{bmatrix}
phrase~\&~nonreferential \\
\text{CONTENT} & \boxed{1} \\
\text{STORE} & \left\{\boxed{1}\right\} \cup \boxed{2} \\
\text{HEAD-DTR} \mid \text{STORE} & \boxed{2}
\end{bmatrix}
\qquad
\begin{bmatrix}
word~\&~nonreferential \\
\text{CONTENT} & \boxed{1} \\
\text{STORE} & \left\{\boxed{1}\right\}
\end{bmatrix}
$$

We can now see how the C-PARAMS value is built up for a sentence:

(196)

$$
\begin{bmatrix}
\text{PHON} & \langle \text{the, dog, snores} \rangle \\
\text{CONTENT} & \boxed{5}\,S(w) \\
\text{C-PARAMS} & \left\{ \boxed{1}w, \boxed{2}Q, \boxed{3}P, \boxed{4}S \right\}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\textit{qnp \& referential} \\
\text{PHON} & \langle \text{the, dog} \rangle \\
\text{CONTENT} & \boxed{1}\begin{bmatrix} w : w = Q(P) \end{bmatrix} \\
\text{C-PARAMS} & \left\{ \boxed{1}, \boxed{2}, \boxed{3} \right\}
\end{bmatrix}
\qquad
\begin{bmatrix}
\text{PHON} & \langle \text{snores} \rangle \\
\text{CONTENT} & \boxed{5}\,S(w) \\
\text{C-PARAMS} & \left\{ \boxed{4}S : name(S, snore) \right\}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\text{PHON} & \langle \text{the} \rangle \\
\text{CONTENT} & \boxed{2}\begin{bmatrix} Q : Q = the'' \end{bmatrix} \\
\text{C-PARAMS} & \left\{ \boxed{2} \right\}
\end{bmatrix}
\qquad
\begin{bmatrix}
\text{PHON} & \langle \text{dog} \rangle \\
\text{CONTENT} & \boxed{3}\begin{bmatrix} P : name(P, dog) \end{bmatrix} \\
\text{C-PARAMS} & \left\{ \boxed{3} \right\}
\end{bmatrix}
$$

For those NPs with a functional analysis (e.g. attributive definites) a slightly different version of the principle is of course required: the function and argument parameters are treated separately and can be contributed individually to either STORE or C-PARAMS.

(197)

$$
\begin{bmatrix}
\textit{phrase} \\
\text{CAT} \mid \text{HEAD} & \textit{nominal} \\
\text{CONTENT} & \begin{bmatrix} \text{FUNC} & \boxed{1a} \\ \text{ARG} & \boxed{1b} \end{bmatrix} \\
\text{STORE} & \boxed{2} \cup \boxed{3} \\
\text{C-PARAMS} & \left(\left\{ \boxed{1a}, \boxed{1b} \right\} - \boxed{2} \right) \cup \boxed{4} \cup \ldots \cup \boxed{n} \\
\text{HEAD-DTR} & \begin{bmatrix} \text{STORE} & \boxed{3} \end{bmatrix} \\
\text{DTRS} & \left\langle \begin{bmatrix} \text{C-PARAMS} & \boxed{4} \end{bmatrix}, \ldots, \begin{bmatrix} \text{C-PARAMS} & \boxed{n} \end{bmatrix} \right\rangle
\end{bmatrix}
$$
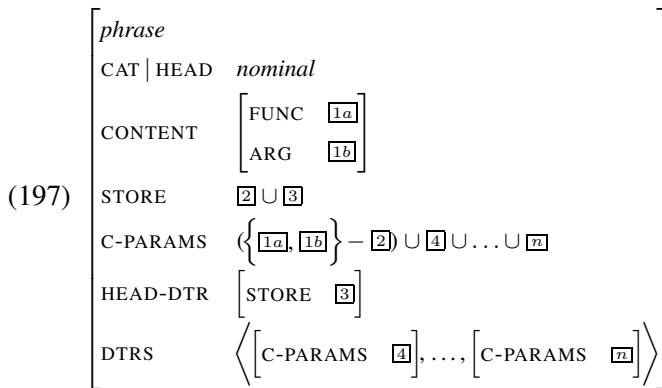
4.4.6   Summary

This section has shown that no NP reprises appear to query a generalized quantifier or property-of-properties, but that reprises of definite NPs can query an individual (or set of individuals), and that this may also be true for certain referential uses of other QNPs.

We have seen that the reprise content hypothesis can be held to in its strong version if a semantic representation of QNPs as denoting witness sets is used. This leads to a relatively simple flat representation, with similarities to that of Hobbs (1983) or the choice function/epsilon term approach. A standard GQ representation can only hold to the weak version of the

hypothesis, making it difficult to explain why reprises do not appear to be able to query GQs.

The next section shows how this can be extended to cope with important issues we have so far only mentioned briefly: quantification, relative quantifier scope, anaphora, non-MON↑ quantifiers and sub-constituent focussing. After that, section 4.6 takes a quick look at some other phrase types, including examining the implications of this QNP analysis for the semantics of determiners.

4.5   Further Issues

4.5.1   Quantification and Scope

So far we have described indefinites and other quantified NPs as denoting sets that are existentially quantified via membership of the STORE feature. This section defines this approach to quantification properly and explains how relative scope between the sets can be expressed given a flat witness-set-based representation.

Quantification, Storage and Retrieval

The approach proposed so far represents all non-definites as existentially quantified sets, and therefore requires a mechanism for introducing this quantification into the semantic content of the sentence at the appropriate level. This can be achieved through the familiar storage method of (Cooper, 1983), using the feature STORE to which existentially quantified elements are added by lexical/phrasal constituents and from which they are retrieved to form part of the sentence semantics.

The lexically-based retrieval mechanism defined by G&S can be used, whereby inherited STORE values are (by default) allowed to be discharged into the QUANTS feature by lexical heads, although a couple of simplifications can be made. As only simultaneous existential quantification is being used (see below), the order of quantifiers is not important – we can therefore represent QUANTS as a *set* rather than a *list*, thus no longer requiring G&S's `order` operator. Both the STORE and QUANTS features can also be treated as sets of *parameters* rather than *quantifiers* (which also turns out to be useful for a treatment of anaphora – see section 4.5.2 below). Our version of the STORE Amalgamation Constraint therefore appears as in AVM (198): a head word's STORE value is defined as the union of the STORE values of its sisters (the members of its ARG-ST list) minus whatever elements are made members of QUANTS . Both QUANTS and STORE are then inherited by a mother by the default Generalised Head Feature Principle (see G&S, p.208). Finally, top-level root clauses are constrained to

have empty STORE values, thus ensuring that all stored parameters have been discharged.

$$
(198) \quad
\begin{bmatrix}
word \\
\text{CONTENT} & \begin{bmatrix} \text{QUANTS} & \boxed{2} \end{bmatrix} \\
\text{STORE} & \{\boxed{1} \cup \ldots \cup \boxed{m}\} - \boxed{2} \\
\text{ARG-ST} & \left\langle \begin{bmatrix} \text{STORE} & \boxed{1} \end{bmatrix}, \ldots, \begin{bmatrix} \text{STORE} & \boxed{m} \end{bmatrix} \right\rangle
\end{bmatrix}
$$

The members of the QUANTS set are now taken to be *simultaneously* quantified over, following Cooper (1993)'s definition of simultaneous quantification for his situation-theoretic reconstruction of DRT (Kamp and Reyle, 1993). A quantified object is viewed as a simultaneous abstract, with the QUANTS set abstracted from the body. Truth conditions are then dependent on finding some appropriate assignment for that abstract – one which assigns values to the members of the abstracted set such that the standard truth conditions hold for the body. More formally, a proposition of the form:

$$
(199) \quad
\begin{bmatrix}
proposition \\
\text{SIT} & s \\
\text{SOA} & \begin{bmatrix} soa \\ \text{QUANTS} & Q \\ \text{NUCL} & \sigma \end{bmatrix}
\end{bmatrix}
$$

is taken as an abstract $\lambda Q.\sigma$, and is true if there exists some assignment f appropriate for $\lambda Q.\sigma$ such that s supports σ under that assignment (written $\sigma\{f\}$). An assignment f is some function that maps the indices in Q onto a particular set of individuals: applying this same mapping to the indices in σ results in $\sigma\{f\}$, which is identical to σ modulo that mapping. If $\sigma\{f\}$ is supported by the situation s for some f, then the existentially quantified proposition is true.

We can now see how a sentence can be built up with a combination of definites and indefinites, contributing respectively to C-PARAMS and STORE/QUANTS (leaving out sub-constituent C-PARAMS now for clarity):[56]

---

[56]In fact, AVM (200) also leaves out C-PARAMS associated with the verb *like* – see section 4.6.3 below.

(200)

$$
\begin{bmatrix}
\text{PHON} & \left\langle \text{the, dog, likes, a, cat} \right\rangle \\
\text{CONTENT} & \boxed{1}\begin{bmatrix} \text{QUANTS} & \boxed{5}\{c\} \\ \text{NUCL} & like(d,c) \end{bmatrix} \\
\text{STORE} & \boxed{6}\{\} \\
\text{C-PARAMS} & \{\boxed{2}d\}
\end{bmatrix}
$$

$$
\boxed{8}\begin{bmatrix}
\textit{qnp \& definite} \\
\text{PHON} & \left\langle \text{the, dog} \right\rangle \\
\text{CONTENT} & \boxed{2}\begin{bmatrix} d:d = Q(P) \end{bmatrix} \\
\text{STORE} & \boxed{7}\{\} \\
\text{C-PARAMS} & \{\boxed{2}\}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\text{PHON} & \left\langle \text{likes, a, cat} \right\rangle \\
\text{CONTENT} & \boxed{1}\begin{bmatrix} \text{QUANTS} & \boxed{5} \end{bmatrix} \\
\text{STORE} & \boxed{6} \\
\text{C-PARAMS} & \{\}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\text{PHON} & \left\langle \text{likes} \right\rangle \\
\text{CONTENT} & \boxed{1}\begin{bmatrix} \text{QUANTS} & \boxed{5} \end{bmatrix} \\
\text{STORE} & \boxed{6} = \boxed{7} \cup \boxed{4} - \boxed{5} \\
\text{ARG-ST} & \left\langle \boxed{8}\begin{bmatrix} \text{STORE} & \boxed{7} \end{bmatrix}, \boxed{9}\begin{bmatrix} \text{STORE} & \boxed{4} \end{bmatrix} \right\rangle
\end{bmatrix}
$$

$$
\boxed{9}\begin{bmatrix}
\textit{qnp \& indefinite} \\
\text{PHON} & \left\langle \text{a, cat} \right\rangle \\
\text{CONTENT} & \boxed{3}\begin{bmatrix} c:c = Q'(P') \end{bmatrix} \\
\text{STORE} & \boxed{4}\{\boxed{3}c\} \\
\text{C-PARAMS} & \{\}
\end{bmatrix}
$$

Representation of Scope

A representation of NPs as denoting witness sets also needs a way of expressing relative scope between the sets introduced by a sentence, both those sets associated with definites that will be fixed in context, and those associated with non-definites which are existentially quantified over. A standard approach of ordering quantifiers cannot apply; instead, relative scope can be expressed by regarding the sets as functionally dependent on one another.

As we already have a functional representation of NPs (motivated by non-referential definites and outlined in section 4.4.1), all that is required is to allow them to take the sets denoted by other NPs as arguments: narrow-scoping NPs will be functional on other wider-scoping sets. The alternative readings of *"every dog_d likes a cat_c"* can then be produced by representing *a cat* either as a simple existentially quantified individual c, or as a functional one $f(d)$, dependent on the set of dogs d via an existentially quantified function f.

This kind of dependence has its precedents: it results in an analysis similar to choice function/epsilon term analyses, and in particular the analysis of von Heusinger (2000, 2002). He represents definites and indefinites as epsilon terms, which are semantically interpreted via choice functions which must be dependent on other indices in the discourse. For narrow-

scoping indefinites, the choice function will be dependent on the index of another wider-scoping epsilon term; for wide-scoping equivalents, the choice function will be dependent only on some fixed referential expression such as the current speaker or temporal index. There are two main differences in the approach proposed here. Firstly, while von Heusinger takes *all* NPs to be dependent on other indices, the current proposal only requires this for those NPs for which a functional analysis seems motivated by the readings of CRs, i.e. non-referential definites, and phrases taking narrow scope – referential definites seem better treated as non-functional, requiring a referential parameter to be grounded. Secondly, the function proposed here is not strictly a choice function, although it is closely related to one: von Heusinger's choice functions take a set as argument (e.g. the set of all cats), and return an element from that set (a cat). For a definite, it is the identity of the choice function which must be con-textually fixed to determine which cat is chosen. In contrast, the function proposed here is that function which takes a situation (or wider-scoping set) as an argument, and returns the definite cat given that argument. This difference is important when considering reprises of non-referential definites: identification of this function does not uniquely identify the referent cat (the argument is needed too), and so a reprise need not query the referent cat's identity, whereas identifying a von Heusinger-style choice function would seem to fix the referent directly.

There are also similarities to the approach of Farkas (1997), in which scope differences are expressed by different assignment functions. In this case the details are less similar to the currently proposed approach, and seem even less suited to explicating reprise readings — a variable associated with a narrow-scoping quantified phrase is evaluated with respect to an assignment function which is dependent on the wider-scoping quantifier; a wide-scoping variable is evaluated with respect to an independent base assignment function — but the basic approach is still one of functional dependence rather than e.g. quantifier raising or movement.

Assigning Scope   This function f remains a member of C-PARAMS or STORE depending on (in)definiteness, according to the Definiteness Principle described in section 4.4.5. The argument d must be identified with the relevant wide-scoping set: where the wide-scope NP is definite and its content is in C-PARAMS, this is achieved by making the narrow-scope argument a member of C-PARAMS and identifying the two during grounding; where the wide-scope NP is indefinite and its content in STORE, it occurs through the anaphoric binding mechanism described in section 4.5.2 below. AVM (201) illustrates the former alternative – a version with a narrow-scope functional indefinite, whose argument will be resolved during C-PARAMS instantiation (for the reading where *a cat* takes narrow scope relative to *the dogs*, identifying the set of dogs d_1 with the functional argument d_2).

(201)

$$
\begin{bmatrix}
\text{PHON} & \langle \text{the, dogs, like, a, cat} \rangle \\[4pt]
\text{CONTENT} & \boxed{1}\begin{bmatrix} \text{QUANTS} & \boxed{5}\{c\} \\ \text{NUCL} & like(d_1, c(d_2)) \end{bmatrix} \\[8pt]
\text{STORE} & \boxed{6}\{\} \\[4pt]
\text{C-PARAMS} & \{\boxed{2}d_1, \boxed{4}d_2\}
\end{bmatrix}
$$

Left daughter:

$$
\boxed{8}\begin{bmatrix}
\textit{qnp \& definite} \\
\text{PHON} & \langle \text{the, dogs} \rangle \\
\text{CONTENT} & \boxed{2}\big[d_1 : d_1 = Q(P)\big] \\
\text{STORE} & \boxed{7}\{\} \\
\text{C-PARAMS} & \{\boxed{2}\}
\end{bmatrix}
$$

Right daughter:

$$
\begin{bmatrix}
\text{PHON} & \langle \text{like, a, cat} \rangle \\
\text{CONTENT} & \boxed{1}\big[\text{QUANTS} \ \boxed{5}\big] \\
\text{STORE} & \boxed{6} \\
\text{C-PARAMS} & \{\boxed{4}\}
\end{bmatrix}
$$

Right daughter's daughters:

$$
\begin{bmatrix}
\text{PHON} & \langle \text{like} \rangle \\
\text{CONTENT} & \boxed{1}\big[\text{QUANTS} \ \boxed{5}\big] \\
\text{STORE} & \boxed{6} = \boxed{7}\cup\boxed{3}-\boxed{5} \\
\text{ARG-ST} & \left\langle \boxed{8}\big[\text{STORE} \ \boxed{7}\big], \boxed{9}\big[\text{STORE} \ \boxed{3}\big] \right\rangle
\end{bmatrix}
$$

$$
\boxed{9}\begin{bmatrix}
\textit{qnp \& indefinite} \\
\text{PHON} & \langle \text{a, cat} \rangle \\
\text{CONTENT} & \begin{bmatrix} c(d_2) : c = Q'(P'), \\ d_2 \subseteq DOM(c) \end{bmatrix} \\
\text{STORE} & \boxed{3}\{c\} \\
\text{C-PARAMS} & \{\boxed{4}d_2\}
\end{bmatrix}
$$

As already pointed out in section 4.4.1, if we want to remove the ambiguity that this introduces (the alternative representations as functional or non-functional) we could take von Heusinger (2002)'s approach of regarding *all* NPs as functional, with the widest scoping elements as functional on some index external to the sentence, e.g. the speaker or utterance situation (the argument will therefore be a member of C-PARAMS). This wouldn't actually give us less work to do overall (we still have to identify the argument in context when grounding) but would remove the ambiguity of representation (leaving us with an ambiguity of reference). However, as already observed in this section, it seems difficult to square such an approach with the apparent meaning of referential reprises, which really do seem to query a simple non-functional referent.

There is a further possible ambiguity in that we have postulated functional NPs with two types of argument – those functional on situations (as for attributive definites) and those functional on other NP witness sets (as for narrow scope here). A simpler view with only situations as arguments might be possible: in the case of narrow-scoping elements, the argument would be a situation linked to another NP, directly analogous to Cooper (1995)'s *individual* situation (a situation for each member of the witness set, which supports the proposition expressed by the sentence for that member). The cost of this view would be that sets of individual situations

must be provided in C-PARAMS/STORE, either by NPs themselves or by verbal predicates. As there is no direct evidence for this, we leave it aside for now as a possible alternative.

Reprises and Scope

This analysis would imply that directly referential reprises can only make any sense when reprising a QNP with widest scope, while reprises of narrow-scoping elements will be read as functional – attempting to identify the function or its argument. While finding corpus examples of multiple-quantifier sentences in dialogue with determinable scope ordering and followed by reprise questions seems to be too much to hope for, invented examples such as example (202) should be paraphrasable as shown, and this seems to be about right:

(202)

A: Every professor relies on their teaching assistant.
B: **Their teaching assistant**?

$\rightsquigarrow$ *"What situation are you intending me to interpret 'their teaching assistant' relative to?"*
$\rightsquigarrow$ *"What are you intending 'their teaching assistant' to refer to for each professor?"*
$\rightsquigarrow$ #*"Which actual person are you referring to by 'their teaching assistant'?"*

4.5.2   Anaphora

Intersentential Anaphora

An account of anaphora seems to follow simply, whereby anaphoric terms such as pronouns are treated like definites – they have referential C-PARAMS whose reference must be established during the grounding process. The constraints on this identification may be slightly different to those for definites: rather than having to identify a referent in the general context, truly anaphoric uses must have to refer to entities already established in the discourse. Deictic uses can be accounted for by assuming that salient referents are introduced into the discourse (or the general context) by external cognitive means.

Details will depend on the model of context being used, and in particular the notion of salience or discourse structure. Whatever the model of context, though, the treatment of NPs as denoting witness sets rather than GQs seems attractive from the point of view of anaphora, as it allows these sets to provide potential referents for anaphors in future utterances. Where these antecedent sets are associated with definites, it is clear that they are already in the context: for indefinites, a protocol will be required to account for their addition thereto.[57]

---

[57]This cannot be as simple as adding an utterance's existentially quantified sets to a discourse record on acceptance: Ginzburg (2001) gives examples of anaphora to entities from unaccepted assertions and even from ungrounded utterances. One way to take these into account might be to allow for the possibility of pronouns which are functional on (sub-)utterances themselves (or, as Ginzburg suggests, utterance situations).

An exception, however, is the quantifier *every*. Individual terms introduced by singular NPs will clearly license singular anaphora; sets of terms associated with their plural counterparts and QNPs with quantifiers such as *all* and *most* will correspondingly licence only plural anaphora. In contrast, *every* also licenses singular anaphora. If we assume that an *every*-QNP denotes a set, it is not clear how a singular individual is provided for reference. If instead we view a singular pronoun as functional on a set, it is not clear why this is not possible for other plural quantifiers.

Intrasentential Anaphora

Accounting for intrasentential anaphora requires a further step. If pronouns (and anaphoric definites) are taken as referring to existentially quantified elements within the same sentence, they can no longer have a contextual parameter associated with them: they do not refer to an element in the context external to the utterance.

It must therefore be the case that elements of C-PARAMS can be removed if they can be identified with an element of QUANTS – i.e. a binding mechanism similar to Poesio (1994)'s *parameter anchoring* and van der Sandt (1992)'s *presupposition binding* (hence the advantage of the implementation of STORE/QUANTS as parameters rather than quantifiers). This mechanism is implemented via a new feature B(OUND)-PARAMS: referential parameters can be members of either C-PARAMS or B-PARAMS, but membership of B-PARAMS is limited to those parameters which can be identified with existentially quantified parameters (i.e. members of STORE/QUANTS). This leads us to the final version of the Definiteness Principle:

$$
(203) \quad
\begin{bmatrix}
word \\
\text{CONTENT} & \boxed{1} \\
\text{STORE} & \boxed{2} \\
\text{C-PARAMS} & \boxed{3} \\
\text{B-PARAMS} & \left\{\boxed{1}\right\} - \boxed{2} - \boxed{3}
\end{bmatrix}
$$

while the restriction on B-PARAMS membership is expressed through the final version of the lexical quantifier storage mechanism:

$$
(204) \quad
\begin{bmatrix}
word \\
\text{CONTENT} & \begin{bmatrix} \text{QUANTS} & \boxed{Q} \end{bmatrix} \\
\text{STORE} & \boxed{S} = \{\boxed{1_a} \cup \ldots \cup \boxed{n_a}\} - \boxed{Q} \\
\text{B-PARAMS} & \{\boxed{1_b} \cup \ldots \cup \boxed{n_b}\} - subset(\boxed{Q} \cup \boxed{S}) \\
\text{ARG-ST} & \left\langle \begin{bmatrix} \text{STORE} & \boxed{1_a} \\ \text{B-PARAMS} & \boxed{1_b} \end{bmatrix}, \ldots, \begin{bmatrix} \text{STORE} & \boxed{n_a} \\ \text{B-PARAMS} & \boxed{n_b} \end{bmatrix} \right\rangle
\end{bmatrix}
$$

To ensure that all members of B-PARAMS are thus discharged, all that is required is to specify top-level sentences (following the conventions of G&S, signs of type *root-cl*) as having empty B-PARAMS. Note that this mechanism can also apply to the arguments of narrow-

scope functional NPs, thus allowing them to be functional from wider-scoping existentially quantified sets. This includes situational arguments, allowing the argument of an attributive definite to be taken as the situation introduced in the utterance (the described situation).

4.5.3   Monotone Decreasing Quantifiers

As mentioned in section 4.4.4 above, B&C point out that it is not sufficient with monotone decreasing (MON↓) cases to show that a predicate holds of a witness set: instead we must show that the witness set contains *all* members of the restriction set of which the predicate holds.

(205) $\exists w[(X \cap A) \subseteq w] \quad \leftrightarrow \quad X \in D(A)$

This means that the representation of QNPs as denoting witness sets proposed here fails to encapsulate the meaning of MON↓ quantifiers (or non-monotone quantifiers such as *exactly two*). The sentence *"Few dogs snore"* does not only convey the fact that the property of snoring holds of some set w containing few dogs (as our simple representation would – see (206)), but also that the property does not hold of any dogs not in w (e.g. as in (207)):

(206) $few'(w, P) \wedge dog(P) \wedge snore(w)$

(207) $few'(w, P) \wedge dog(P) \wedge snore(w) \wedge \neg\exists w'[(w' \subseteq P) \wedge (w \subset w') \wedge snore(w')]$

One solution might be to appeal to pragmatics: Hobbs (1996) solves the problem by use of a pragmatic constraint which strengthens the sentence meaning accordingly: *few dogs snore* is taken just as the assertion that there is a set containing few dogs, all of whom snore, but this is strengthened by an abductive process to the assertion that this set is the maximal set of snoring dogs.

Another would of course be to regard the content of QNPs as GQs rather than witness sets, but of course this means only the weak hypothesis can hold (see above).

Another possibility is the view of MON↓ quantifiers as the negation of their MON↑ counterparts (*few dogs snore* is truth-conditionally equivalent to *most dogs don't snore*). This has been much explored in the DPL tradition of GQs (see e.g. van den Berg, 1996).

Complement Set Anaphora

One of the advantages of this last approach is that it allows for an explanation of the phenomenon of *complement set anaphora* noticed by Moxey and Sanford (1987, 1993). Kibble (1997a,b) sees sentences with such quantifiers as ambiguous between internal and external negation (*most dogs don't snore* vs. *it's not true that most dogs snore*), giving rise to the possibility of complement set (the dogs who don't snore) and reference set (the dogs who do) anaphora respectively.

An interesting question is therefore whether reprise questions of MON↓ QNPs can query the reference or complement set. The pragmatic approach would suggest only the reference set is possible, the negation approach the reverse. Sadly, corpus examples of MON↓ QNP reprises are rare: most seem to be best paraphrased as sub-constituent readings, querying either the CN predicate or the logical quantifier relation:

(208)[58]

Lorna:	Oh shit! I've gotta ring mum. Tell mum no meat.
Kathleen:	**No meat?**
Lorna:	I'm not allowed to get meat and stuff.
Kathleen:	Why?
Lorna:	Cos we're vegetarians!

$\rightsquigarrow$ *"Is it really **meat**$_P$ you're saying to tell mum no P?"*
$\rightsquigarrow$ *"Is it really **no**$_N$ you're saying to tell mum N meat?"*

(209)[59]

Merielle:	they wanted it early, I don't want anything!
Harold:	Do you want crisps?
Merielle:	Nothing!
Harold:	**Nothing?**
	Ooh, okay!
Martine:	I'll have a <pause> wine thanks?

$\rightsquigarrow$ *"Is it really **no**$_N$ you're saying you want N things?"*

But some do seem to allow for reference set reference, and possibly for complement set reference as well, although this seems less clear:

(210)[60]

Anon 1:	Did any of them the lads that you the men that you went away with. Did they come back?
Richard:	Not all.
Anon 1:	**Not all of them?**
Richard:	Oh no.
Anon 1:	Were any of them.

$\rightsquigarrow$ *"Who are you telling me did come back?"*
$\rightsquigarrow$?*"Who are you telling me didn't come back?"*

Kibble gives this example of complement set anaphora:

(211) | BBC News: Not all of the journalists agreed, among them the BBC's John Simpson.

where *them* is construed to refer to the group of journalists who did *not* agree.[61] An

---

[58]BNC file KCW, sentences 2204–2210
[59]BNC file KD8, sentences 1371–1376
[60]BNC file HEU, sentences 360–365
[61]Arguments have been made (e.g. Corblin, 1996) for regarding such anaphora as not referring to the com-

imagined reprise version seems easier to construe as querying the complement set:

(212)

A: Not all of the journalists agreed.

B: **Not all of them?**

A: John Simpson was pretty combative.
    Marr and Paxman didn't like it much either.

$\leadsto$ *"Who do you mean **didn't** agree?"*

If so, a more consistent approach would be to view MON↓ QNPs as denoting *pairs* of reference and complement sets $\langle R, C \rangle$. The reference set R is, as with MON↑ QNPs, a witness set; the complement set C is $(A - R)$ (for a quantifier living on A). Such a pair might be paraphrased as *"R as opposed to C"*, and can be interpreted as follows:

(213) $snore(\langle R, C \rangle) \quad \leftrightarrow \quad (R \subseteq snore) \wedge (C \cap snore = \emptyset)$

An corresponding HPSG analysis can be constructed as shown in AVM (214) (with the λ-abstract equivalent shown in (215)).

(214)

$$\begin{bmatrix} \text{PHON} & \langle \text{few, dogs} \rangle \\ \text{CONTENT} & \begin{bmatrix} \text{REF} & \boxed{1}[r:r = Q(P)] \\ \text{COMP} & \boxed{2}[c:c = (P-r)] \end{bmatrix} \\ \text{C-PARAMS} & \boxed{5} \cup \boxed{6} \\ \text{STORE} & \{\boxed{1}, \boxed{2}\} \\ \text{DTRS} & \left\langle \begin{bmatrix} det \\ \text{CONTENT} & \boxed{3}[Q:Q = few''] \\ \text{C-PARAMS} & \boxed{5}\{\boxed{3}\} \end{bmatrix}, \begin{bmatrix} nominal \\ \text{CONTENT} & \boxed{4}[P:name(P,dog)] \\ \text{C-PARAMS} & \boxed{6}\{\boxed{4}\} \end{bmatrix} \right\rangle \end{bmatrix}$$

(215) $\lambda\{Q, P\}[Q = few'', name(P, dog)].\exists\{r, c\}[r = Q(P), c = (P-r)].\langle r, c \rangle$

Most such QNPs (as with most other QNPs) will presumably be non-referential and thus will not contribute to C-PARAMS, with the pair of sets instead existentially quantified via STORE, as shown in (215) above. What is contributed in any referential cases depends on whether we believe in complement set reprises – if so, the pair $\langle R, C \rangle$ will be made a member of C-PARAMS, thus holding to the strong hypothesis as in AVM (216); if not, just R, as in AVM (217).

---

plement set but rather as modified reference to the maximal set (in example (211), a generalised reference to all journalists). See (Nouwen, 2003) for some arguments against this view and in favour of genuine complement set reference.

$$(216) \quad \begin{bmatrix} \text{CONTENT} & \begin{bmatrix} \text{REF} & \boxed{1}\big[r:r=Q(P)\big] \\ \text{COMP} & \boxed{2}\big[c:c=(P-r)\big] \end{bmatrix} \\ \text{C-PARAMS} & \big\{\boxed{1},\boxed{2}\big\}\cup\boxed{5}\cup\boxed{6} \\ \text{STORE} & \{\} \end{bmatrix} \quad \lambda\big\{r,c,Q,P\big\}[\ldots].\langle r,c\rangle$$

$$(217) \quad \begin{bmatrix} \text{CONTENT} & \begin{bmatrix} \text{REF} & \boxed{1}\big[r:r=Q(P)\big] \\ \text{COMP} & \boxed{2}\big[c:c=(P-r)\big] \end{bmatrix} \\ \text{C-PARAMS} & \big\{\boxed{1}\big\}\cup\boxed{5}\cup\boxed{6} \\ \text{STORE} & \big\{\boxed{2}\big\} \end{bmatrix} \quad \lambda\big\{r,Q,P\big\}[\ldots].\exists\big\{c\big\}[\ldots].\langle r,c\rangle$$

The existence of both members of the pair now helps explain why they are both possible anaphoric referents: and so why (only) MON↓ QNPs license complement-set reference. As it stands, this says nothing about the relative preference for reference set anaphora observed by Nouwen (2003), or the possibility that not all MON↓ quantifiers license complement set anaphora that he also raises – for example, numerical decreasing quantifiers such as *"fewer than three"* do not seem to license complement set anaphora, while proportional decreasing versions such as *"less than half"* do. Nouwen's proposed solution centres around a treatment of anaphora whereby only those sets that can be semantically inferred to be non-empty can be considered as possible referents: *"fewer than three dogs snore"* does not entail that there are any non-snoring dogs (and so the complement set may be empty), whereas *"less than half of the dogs snore"* does (and so the complement set must be non-empty and supports reference). This seems perfectly applicable to the approach proposed here, and further investigation of MON↓ reprises (particularly if more data can be obtained) may help determine its empirical suitability.

Note that such a treatment requires the possibility that the complement set c be the empty set; the same must also be true for the reference set r, so that sentences such as *"few dogs snore"* and *"no dogs snore"* have the correct truth conditions – both are true if no dogs snore. This will be taken up in more detail in section 5.3.3 when discussing sluices.

4.5.4   Sub-Constituent Focussing

The inheritance of C-PARAMS from daughters defined in section 4.4.5 goes some way towards accounting for the sub-constituent readings that always seem available (especially when a constituent is intonationally stressed), but we also require an explanation of how the sub-constituent becomes focussed in order to assign the relevant content to the reprise question. This is sketched out relatively briefly here.

Engdahl and Vallduví (1996)'s analysis of information structure in HPSG is assumed, with a feature INFO-STRUCT divided into FOCUS and GROUND, with the contents of each

linked (in English at least) to intonation. Reprise questions are now taken to be querying the FOCUSsed component (and checking that the GROUND components are indeed given in context by the utterance being clarified).[62]

To achieve this we use (Engdahl et al., 1999; Ginzburg, forthcoming)'s requirement that an utterance with a given FOCUS/GROUND partition requires for its felicity a MAX-QUD question whose abstracted parameter set corresponds to the FOCUSsed constituents. To take Engdahl et al. (1999)'s example, *"JILL likes Bill"*, where *JILL* is focussed, requires the question $?x.like(x, b)$ (or *"Who likes Bill?"*) to be under discussion; *"Jill likes BILL"* conversely requires $?x.like(j, x)$ (or *"Who does Jill like?"*).

We can define a focussed version of G&C's standard *headed-fragment-phrase*, the type used for clausal reprise fragments, which expresses this requirement. The standard version (shown simplified in AVM (218) requires that a bare fragment be syntactically parallel to the contextually salient utterance SAL-UTT, and be co-referential with it (thus giving rise to the correct overall content via the MAX-QUD – see sections 2.3.5 and 2.5.2).

$$(218) \quad \begin{bmatrix} \textit{hd-frag-ph} \\ \text{HEAD-DTR} \quad \begin{bmatrix} \text{CAT} & \boxed{C} \\ \text{CONT} \,|\, \text{INDEX} & \boxed{I} \end{bmatrix} \\ \text{CTXT} \,|\, \text{SAL-UTT} \quad \begin{bmatrix} \text{CAT} & \boxed{C} \\ \text{CONT} \,|\, \text{INDEX} & \boxed{I} \end{bmatrix} \end{bmatrix}$$

The focussed version must ensure that not only is this the case, but that the FOCUSsed constituent show the same parallelism and co-reference with a sub-constituent of SAL-UTT which is also the part abstracted to make the MAX-QUD question. The GROUND part is taken to have a similar requirement of parallelism with some other sub-constituent – it may be that a stronger constraint is possible and that phonological identity is in fact required, but this

---

[62]This does not entirely explain why the GROUND components are present in the reprise at all: presumably this is either to help disambiguate the exact source constituent being clarified, or just to make the reprise more syntactically palatable.

suffices for the current purposes.

$$(219) \quad \begin{bmatrix} \textit{focus-hd-frag-ph} \\[4pt] \text{HEAD-DTR} \quad \begin{bmatrix} \text{CAT} & \boxed{C} \\ \text{CONT} \,|\, \text{INDEX} & \boxed{I} \end{bmatrix} \\[10pt] \text{INFO-STRUCT} \quad \begin{bmatrix} \text{FOCUS} & \begin{bmatrix} \text{CAT} & \boxed{C_f} \\ \text{CONT} \,|\, \text{INDEX} & \boxed{I_f} \end{bmatrix} \\[10pt] \text{GROUND} \,|\, \text{LINK} & \begin{bmatrix} \text{CAT} & \boxed{C_g} \\ \text{CONT} \,|\, \text{INDEX} & \boxed{I_g} \end{bmatrix} \end{bmatrix} \\[20pt] \text{CTXT} \quad \begin{bmatrix} \text{SAL-UTT} \quad \begin{bmatrix} \text{CAT} & \boxed{C} \\ \text{CONT} \,|\, \text{INDEX} & \boxed{I} \\[6pt] \text{CONSTITS} & \left\{ \ldots \begin{bmatrix} \text{CAT} & \boxed{C_f} \\ \text{CONT} & \boxed{1}\begin{bmatrix} \text{INDEX} & \boxed{I_f} \end{bmatrix} \end{bmatrix} \ldots \begin{bmatrix} \text{CAT} & \boxed{C_g} \\ \text{CONT} \,|\, \text{INDEX} & \boxed{I_g} \end{bmatrix} \ldots \right\} \end{bmatrix} \\[10pt] \text{MAX-QUD} \quad \begin{bmatrix} \text{PARAMS} & \left\{ \boxed{1} \right\} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The clausal CR question coercion operation now must allow for a MAX-QUD question to be created which asks about only a sub-constituent of the SAL-UTT, as long as it is that sub-constituent that has contributed the C-PARAM that is being queried. This is simply stated as follows:

$$(220) \quad \begin{bmatrix} \text{C-PARAMS} & \left\{ \ldots, \boxed{1}, \ldots \right\} \\[6pt] \text{CONSTITS} & \left\{ \ldots, \boxed{2}\begin{bmatrix} \text{CONSTITS} & \left\{ \ldots, \begin{bmatrix} \text{CONTENT} & \boxed{1} \end{bmatrix}, \ldots \right\} \end{bmatrix}, \ldots \right\} \\[6pt] \text{CONTENT} & \boxed{4} \end{bmatrix} \qquad \text{(original utterance)}$$

$$\Rightarrow$$

$$\begin{bmatrix} \text{CONTEXT} & \begin{bmatrix} \text{SAL-UTT} & \boxed{2} \\ \text{MAX-QUD} & ?\boxed{1}.\boxed{4} \end{bmatrix} \end{bmatrix} \qquad \text{(partial reprise context description)}$$

This operation produces a context that when combined with the constraint in AVM (219) and the standard method of deriving sentential content from MAX-QUD, gives the required result: the overall content of the fragment is the question of whether a particular assertion was made, and the contextual MAX-QUD question is one with only the focussed sub-constituent in

the abstracted set.

(221)

$$
\begin{bmatrix}
\text{PHON} & \left\langle \text{the, DOG} \right\rangle \\[4pt]
\text{CONT} & \begin{bmatrix} ?.\boxed{4} \end{bmatrix} \\[4pt]
\text{HEAD-DTR} & \begin{bmatrix} \text{CONT} \mid \text{INDEX} & w \end{bmatrix} \\[4pt]
\text{INFO-STRUCT} & \begin{bmatrix}
\text{FOCUS} & \begin{bmatrix} \text{PHON} & \left\langle \text{DOG} \right\rangle \\ \text{CONT} \mid \text{INDEX} & P \end{bmatrix} \\[10pt]
\text{GROUND} \mid \text{LINK} & \begin{bmatrix} \text{PHON} & \left\langle \text{the} \right\rangle \\ \text{CONT} \mid \text{INDEX} & the'' \end{bmatrix}
\end{bmatrix} \\[20pt]
\text{CTXT} & \begin{bmatrix}
\text{SAL-UTT} & \begin{bmatrix}
\text{PHON} & \left\langle \text{the, dog} \right\rangle \\
\text{CONTENT} & \begin{bmatrix} w : w = the''(P) \end{bmatrix} \\
\text{CONSTITS} & \left\{ \begin{array}{l} \ldots \begin{bmatrix} \text{PHON} & \left\langle \text{dog} \right\rangle \\ \text{CONTENT} & \boxed{1}\begin{bmatrix} P : name(P, dog) \end{bmatrix} \end{bmatrix} \ldots \\[14pt] \ldots \begin{bmatrix} \text{PHON} & \left\langle \text{the} \right\rangle \\ \text{CONT} \mid \text{INDEX} & the'' \end{bmatrix} \ldots \end{array} \right\}
\end{bmatrix} \\[40pt]
\text{MAX-QUD} & ?\boxed{1}.\boxed{4}\, assert(a, b, \ldots the''(P) \ldots)
\end{bmatrix}
\end{bmatrix}
$$

For a constituent CR equivalent, the same overall method can be used, associating the focussed part with the abstracted set of MAX-QUD, with slight differences in the definitions caused by the utterance-anaphoric nature of constituent CRs. The normal utterance-anaphoric phrase type, originally introduced in section 2.3.5 and shown slightly simplified in AVM (222), is defined to denote the SAL-UTT sign and is constrained to be phonologically parallel to it. As shown in AVM (223), a focussed version of this phrase type can now be defined which interacts with information structure as desired: while content is still the same, the focus and ground parts must also denote and be phonologically parallel to sub-constituents of

SAL-UTT, and the focussed part must correspond to the abstracted parameter of MAX-QUD:

$$(222) \quad \begin{bmatrix} \textit{utt-anaph-ph} \\ \text{PHON} & \boxed{P} \\ \text{CONT} & \boxed{1} \\ \text{CTXT} \mid \text{SAL-UTT} & \boxed{1}\begin{bmatrix} \text{PHON} & \boxed{P} \end{bmatrix} \end{bmatrix}$$

$$(223) \quad \begin{bmatrix} \textit{focus-utt-anaph-ph} \\ \text{PHON} \;\; \boxed{P} \\ \text{CONT} \;\; \boxed{1} \\ \text{INFO-STRUCT} \begin{bmatrix} \text{FOCUS} & \begin{bmatrix} \text{PHON} & \boxed{P_f} \\ \text{CONT} & \boxed{2} \end{bmatrix} \\ \text{GROUND} \mid \text{LINK} & \begin{bmatrix} \text{PHON} & \boxed{P_g} \end{bmatrix} \end{bmatrix} \\ \text{CTXT} \begin{bmatrix} \text{SAL-UTT} \;\; \boxed{1} \begin{bmatrix} \text{PHON} & \boxed{P} \\ \text{CONSTITS} & \left\{ \ldots \boxed{2}\begin{bmatrix} \text{PHON} & \boxed{P_f} \\ \text{CONT} & \boxed{3} \end{bmatrix} \ldots \begin{bmatrix} \text{PHON} & \boxed{P_g} \end{bmatrix} \ldots \right\} \end{bmatrix} \\ \text{MAX-QUD} \begin{bmatrix} \text{PARAMS} & \left\{ \boxed{3} \right\} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Again, the contextual coercion operation for constituent CRs must now be able to produce a contextual question about the focussed constituent rather than the whole SAL-UTT utterance:

$$(224) \quad \begin{bmatrix} \text{C-PARAMS} & \left\{ \ldots, \boxed{1}, \ldots \right\} \\ \text{CONSTITS} & \left\{ \ldots, \boxed{2}\begin{bmatrix} \text{CONSTITS} & \left\{ \ldots, \boxed{3}\begin{bmatrix} \text{CONTENT} & \boxed{1} \end{bmatrix}, \ldots \right\} \end{bmatrix}, \ldots \right\} \end{bmatrix}$$ (original utterance)

$$\Rightarrow$$

$$\begin{bmatrix} \text{CONTEXT} & \begin{bmatrix} \text{SAL-UTT} & \boxed{2} \\ \text{MAX-QUD} & ?\boxed{1}.spkr\_meaning\_rel(a, \boxed{3}, \boxed{1}) \end{bmatrix} \end{bmatrix}$$ (partial reprise context description)

Combining this operation with the new focussed definition of the utterance-anaphoric phrase type will now give the desired reading for a focussed version of a constituent CR (in AVM (225), with the CN focussed). The sentential content is now a question about the

intended content of the focussed sub-constituent:

(225)

$$
\begin{bmatrix}
\textit{constit-clar-int-cl} \\
\text{PHON} \quad \boxed{P}\langle \text{the, DOG} \rangle \\
\text{CONT} \quad \boxed{4} \\[2mm]
\text{INFO-STRUCT} \begin{bmatrix}
\text{FOCUS} \begin{bmatrix} \text{PHON} & \boxed{P_f}\langle \text{DOG} \rangle \\ \text{CONT} & \boxed{3} \end{bmatrix} \\[3mm]
\text{GROUND} \,|\, \text{LINK} \begin{bmatrix} \text{PHON} & \boxed{P_g}\langle \text{the} \rangle \end{bmatrix}
\end{bmatrix} \\[8mm]
\text{CTXT} \begin{bmatrix}
\text{SAL-UTT} \begin{bmatrix}
\text{PHON} & \boxed{P}\langle \text{the, dog} \rangle \\
\text{CONTENT} & \begin{bmatrix} w : w = the''(P) \end{bmatrix} \\[3mm]
\text{CONSTITS} & \left\{ \ldots \boxed{3}\begin{bmatrix} \text{PHON} & \boxed{P_f}\langle \text{dog} \rangle \\ \text{CONTENT} & \boxed{1}\begin{bmatrix} P : name(P, dog) \end{bmatrix} \end{bmatrix} \ldots \right. \\ & \left. \ldots \begin{bmatrix} \text{PHON} & \boxed{P_g} \end{bmatrix} \ldots \right\}
\end{bmatrix} \\[8mm]
\text{MAX-QUD} \quad \boxed{4}\Big[?\boxed{1}.spkr\_meaning\_rel(a, \boxed{3}, \boxed{1}) \Big]
\end{bmatrix}
\end{bmatrix}
$$

For an analysis of *wh*-versions of such reprise questions, all that is required is that the *wh*-phrase itself be focussed, which seems intuitively reasonable (and see Artstein, 2002, for some more formal arguments). The same analysis will then require reprises like *"The/a what?"* (see examples (162) and (169) above) to query a CN sub-constituent rather than the whole NP.

This is by no means a complete analysis of this phenomenon, but does at least show that an approach is possible which fits with standard notions of information structure and context, and gives the required CR meanings. A complete analysis of this phenomenon will require further investigation into what the parallelism requirements for the various parts (particularly the ground components) really are, and may require an account of *focus spreading* from CN to NP: it seems plausible that a reprise even with the CN intonationally focussed may be interpreted as querying the NP referent. This should be possible, again using Engdahl and Vallduví (1996)'s analysis, but the usual assumption that focus spreads from the *most oblique* daughter to the mother does not appear to hold in this case (intuitively at least – as far as I am aware accounts of focus spreading have never considered phenomena at this low a level, within NPs).

Note also that a full account must consider languages such as Hebrew and Romanian, which can express definite descriptions with single (inflected) words. As in the English examples here, (at least) the referent and predicate readings are available from a definite NP reprise[63] – yet this reprise is of a single word. In this case a theory of sub-constituent focussing must allow for morphemes to be focussed, rather than just whole words.

---

[63]Thanks to George & Corina Dindelegan for discussions of Romanian data.

4.6   Other Phrase Types

The analysis has now been extended to cover common nouns, definite & indefinite NPs, pro-nouns, demonstratives and of course proper names. As section 3.3 showed, these are the main classes that constitute sources of CRs, and others are relatively rare (so less important for a grammar to cover). This section now turns to look briefly at some of these other classes: firstly determiners, as their analysis is to a large extent already dictated by the analysis of nouns and NPs; next *wh*-phrases, which will be important for a grammar when constructing CRs themselves, if not as sources; then verbs and a brief look at a general treatment for other content and function words.

4.6.1   Determiners

Where does the analysis so far leave us with regard to determiners? A view of NPs as denoting witness sets and of CNs as denoting predicates (properties of individuals) seems to leave us with a view of determiners as denoting functions from the CN predicates to the NP sets (i.e. functions of type $(e{\rightarrow}t){\rightarrow}e$). In a model-theoretic sense, they would therefore denote relations between two sets (the equality relation for *every*, a relation that picks out an epsilon term for *a/some*, a relation that picks out a set of a particular cardinality for *two/three*).

The alternative view of NPs as denoting GQs, on the other hand, would force us to view determiners as denoting functions from CN predicates to GQs (sets of sets) – essentially the Montagovian view of determiners as functions of type $(e{\rightarrow}t){\rightarrow}((e{\rightarrow}t){\rightarrow}t)$.

Do either of these fit with what determiner reprise questions seem to mean?

Evidence

Determiner-only reprises certainly exist, but seem to be rare: the only suitable examples found through corpus investigation involved numerals (see examples (226) and (227)) – as was the case for most of the determiner sources found in chapter 3.[64]

	Marsha:	yeah that's it, this, she's got three rottweiler's now and
	Sarah:	**three?**
(226)[65]	Marsha:	yeah, one died so only got three now `<laugh>`
	$\leadsto$	*"Is it **three**$_N$ you are saying she's got N rottweilers?"*

---

[64]The only non-numerical determiner-only reprise questions found were *reprise gap* forms: i.e. not actually querying the determiner but rather whatever came after it.

[65]BNC file KP2, sentences 295–297

(227)[66]

Terence:	It's thirty eight overs so I mean they've got another twelve overs yet.
Margaret:	To get about twenty runs.
Terence:	Ten runs oh no less than that, ten runs now.
Margaret:	**Ten?**
Terence:	Well twelve I think, hundred and forty six they've got now.
Margaret:	Oh that would be twelve then. <pause>

⤳   *"Is it **ten**$_N$ you are saying they've got N runs to get?"*

For these examples, the query appears to concern the cardinality of the set under discussion, which fits quite nicely with the idea of determiners as denoting set relations.

For other determiners, we have to rely on intuition (example (229)), and on those QNP reprise examples mentioned in section 4.4 above in which the determiner appears to be stressed, e.g. example (177), repeated here as example (228):

(228)[67]

Richard:	No I'll commute every day
Anon 6:	**Every day?**
Richard:	as if, er Saturday and Sunday
Anon 6:	And all holidays?
Richard:	Yeah <pause>

⤳   *"Is it **every**$_N$ that you are saying you'll commute on N days?"*

(229)[68]

A:	Is that the shark?
B:	**The?**
A:	You don't think there's more than one, surely?

⤳   *"Is it **the unique**$_N$ that you are asking whether that's the N shark?"*

Again, these readings do seem to fit quite nicely with the idea of determiners as denoting set relations, and perhaps less so with that of relations between sets and sets of sets. Note that all of the paraphrases above are clausal in nature – a constituent question *"What do you mean by 'twenty'?"* seems very unlikely, at least for a native speaker.

Another possible reading seems to be one asking about the situation in which the quantifier relation is being used. This could be accounted for in terms of situated relations (functional on situations), analogous to the functional sets discussed briefly in section 4.4.1 and in more detail in section 4.5.1 below.

However, the sparsity of the evidence and the difficulty of pinning down a definitive paraphrase mean it is difficult to make any strong claims here: but we can say that determiner reprises provide no counter-evidence to the analysis of section 4.4.

---

[66]BNC file KE2, sentences 9500–9505
[67]BNC file KSV, sentences 257–261
[68]Adapted from the film "Jaws". In the film, the speaker says: It's a shark, but not **the** shark.

HPSG Analysis

Both views can be easily accommodated within the framework built up so far. If determiners are represented as relations between sets, they are the Q'' relations described in section 4.4.4 above, and a determiner would therefore be represented as in AVM (230):

$$
(230) \quad \begin{bmatrix} \text{PHON} & \langle \text{the} \rangle \\ \text{CONTENT} & \boxed{1}\big[Q{:}Q = the'' \big] \\ \text{C-PARAMS} & \{\boxed{1}\} \end{bmatrix}
$$

where the'' is defined as before as the relation which picks out a witness set given a particular CN predicate. On the QNP-as-GQ view, the representation would look very similar:

$$
(231) \quad \begin{bmatrix} \text{PHON} & \langle \text{the} \rangle \\ \text{CONTENT} & \boxed{1}\big[Q{:}Q = the \big] \\ \text{C-PARAMS} & \{\boxed{1}\} \end{bmatrix}
$$

but here the is the standard function that takes a CN predicate and returns a GQ. In both cases this relation is shown as a member of C-PARAMS, and if we are to account for the evident possibility of determiner reprises (although they are rare) then this must be the case: determiners are therefore seen as having to pick out a relation from context. Presumably such relations are usually familiar to any conversational participant, ruling out constituent questions as a referent will always be found – the clausal reading will occur when the participant finds the apparently intended relation surprising, say, or inconsistent with what is already known. This is reflected in the definitions above, where equality is used (the relation found in context must be identical to a given relation the'') rather than the looser restriction on predicate name used before for nouns.

4.6.2 *WH*-Phrases

Our discussion of NPs did not mention *wh*-phrases. How should their semantic content be represented so as to be consistent with what their reprises seem to mean?

Very few examples of reprises of *"what/which N"* phrases (i.e. those including a CN) were found, so we have also looked at reprises of bare *wh*-words, as shown in the examples below. Examination of reprises of both types suggests that the query can concern a property but not a referent. In *"what/which N"* examples (see examples (232) and (233)) we see the familiar sub-constituent readings (querying the CN predicate or occasionally the determiner

relation), but referent readings seem impossible:

(232)[69]

	Pat:	Is it your tummy?
	Charlotte:	`<unclear>` the blue in it.
	Carole:	Pardon?
	Charlotte:	What blue is it?
	Carole:	**What blue?**
		`<pause>` Well I hope it's not a blue tummy.
	Pat:	`<laugh>`
	$\rightsquigarrow$	*"Is it **blue**$_P$ you are asking about what P it is?"*
	$\rightsquigarrow$	#*"Which blue are you asking which it is?"*

(233)[70]

	Unknown:	How many procedures have we actually audited so far Richard?
	Richard:	**How many procedures?**
	Unknown:	Yeah.
	Unknown:	`<unclear>`.
	Unknown:	No, I know they don't but is there anything that we haven't audited then.
	$\rightsquigarrow$	*"Is it **procedures**$_P$ you are asking about how many Ps?"*
	$\rightsquigarrow$	*"Is it a **number** of procedures you are asking about?"*
	$\rightsquigarrow$	#*"Which procedures are you asking how many of them there are?"*

Bare *wh*-phrase examples seem to query a similar predicate, presumably expressed as part of the lexical semantics of the *wh*-word itself. Referent readings seem impossible in all cases.

(234)[71]

	Charlotte:	Why does the dustman have to take it away?
	Larna:	No not the dustman, the postman
	Charlotte:	Why does the postman have to take all the letters away?
	Larna:	**Why?**
		Well he takes them to the post office
	Charlotte:	Yeah
	Larna:	then the post office sorts them out
	$\rightsquigarrow$	*"Is it a **reason** you are asking for?"*
	$\rightsquigarrow$	#*"Which reason are you asking for?"*

---

[69]BNC file KBH, sentences 3127–3132
[70]BNC file KM4, sentences 920–924
[71]BNC file KD1, sentences 434–440

	Guy:	we're making
	Unknown:	Yes, erm, erm obviously not made it very well, but erm, but the word acceptable ought to be altered to increased `<pause>`
	Guy:	Where?
(235)[72]	Unknown:	**Where?**
	Unknown:	it'll be in the performance er, erm backwards
	Unknown:	This is page four
	$\rightsquigarrow$	*"Is it the **location/page number** you are asking for?"*
	$\rightsquigarrow$	#*"Which location/page are you asking for?"*

	Rose:	So, shall we say Canterbury? When?
	Unknown:	**When?**
	Unknown:	Be about November?
(236)[73]	Unknown:	Early November?
	$\rightsquigarrow$	*"Is it a **time** you are asking for?"*
	$\rightsquigarrow$	#*"Which time are you asking for?"*

In other words, *wh*-phrases seem to have the same clarificatory potential and afford the same reprise readings as standard non-referential indefinites. Given the strong version of our hypothesis, the simplest and most consistent analysis therefore seems to be that these two phrase types have similar semantic representations, in that they both represent terms (or sets of terms) which are not added to C-PARAMS, and thus cannot lead to referent reprise readings. C-PARAMS therefore contains only the contributions from sub-constituents (or for bare *wh*-phrases, lexical semantics), and thus only the sub-constituent readings are available.

In terms of local semantic content, then, indefinites and *wh*-phrases can be given the same representation (a parameter). In terms of sentential semantics, the distinction between the two must come in the fact that the parameters denoted by *wh*-phrases are not existentially quantified but queried: given a view of questions as λ-abstracts, they are part of the abstracted set. In HPSG terms, they must become members of the PARAMS set rather than QUANTS.

This can be achieved with the same general quantifier storage and retrieval mechanism as used for the existentially quantified members of QUANTS (see section 4.5.1 above). *WH*-parameters are added to STORE but given a distinct type to distinguish them from parameters to be existentially quantified: the type *parameter* is split into two subtypes, *wh-param* and *non-wh-param*. *wh*-phrases can then be defined, via the type *wh-phrase*, to have a *wh-param* as their content. By making *wh-phrase* a subtype of *nonreferential* we ensure that this content is contributed to STORE rather than C-PARAMS, explaining the impossibility of referent

---

[72]BNC file KN3, sentences 150–155

[73]BNC file KLS, sentences 1141–1145

reprises, as shown in AVM (237).

$$
(237) \quad
\begin{bmatrix}
\textit{wh-phrase} \\[4pt]
\text{CONTENT} & \boxed{1}\begin{bmatrix}\textit{wh-param}\end{bmatrix} \\[6pt]
\text{STORE} & \left\{\boxed{1}\right\} \cup \boxed{2} \\[6pt]
\text{HEAD-DTR} \mid \text{STORE} & \boxed{2}
\end{bmatrix}
$$

The feature PARAMS is then defined as a set of *wh-param*s, while QUANTS is a set of *non-wh-param*s, preventing the parameters contributed by *wh*-phrases from being discharged into it. G&S's Interrogative Retrieval Constraint (AVM (238) – see G&S, p.227) now takes care of the rest: the *wh*-parameters can only be removed from STORE by becoming members of PARAMS, forcing the content of the sentence to be a question.

$$
(238) \quad
\begin{bmatrix}
\textit{interrogative-clause} \\[4pt]
\text{CONTENT} & \begin{bmatrix}\textit{question} \\ \text{PARAMS} & \boxed{1}\end{bmatrix} \\[8pt]
\text{STORE} & \boxed{2} \\[4pt]
\text{HEAD-DTR} & \begin{bmatrix}\text{STORE} & \boxed{1} \uplus \boxed{2}\end{bmatrix}
\end{bmatrix}
$$

In more detail, AVM (239) now shows how a *wh*-phrase analysis will be built up for a phrase including a CN. As usual, both CN and determiner contribute contextual parameters (corresponding to noun predicate and logical determiner relation respectively), but as the mother's content is a member of STORE, these two become the only members of C-PARAMS.

$$
(239) \quad
\begin{bmatrix}
\textit{wh-phrase} \\[4pt]
\text{PHON} & \langle \text{which, dog} \rangle \\[6pt]
\text{CONTENT} & \boxed{1}\begin{bmatrix}\textit{wh-param} \\ w:w = Q(P)\end{bmatrix} \\[8pt]
\text{STORE} & \left\{\boxed{1}\right\} \cup \boxed{5} \\[6pt]
\text{C-PARAMS} & \boxed{2} \cup \boxed{3} \\[4pt]
\text{HEAD-DTR} & \boxed{4} \\[6pt]
\text{DTRS} & \left\langle
\begin{bmatrix}
\text{PHON} & \langle \text{which} \rangle \\
\text{C-PARAMS} & \boxed{6}\begin{bmatrix}Q:Q = which''\end{bmatrix} \\
\text{C-PARAMS} & \boxed{2}\left\{\boxed{6}\right\}
\end{bmatrix},
\boxed{4}\begin{bmatrix}
\text{PHON} & \langle \text{dog} \rangle \\
\text{STORE} & \boxed{5}\{\} \\
\text{CONTENT} & \boxed{7}\begin{bmatrix}P:name(P,dog)\end{bmatrix} \\
\text{C-PARAMS} & \boxed{3}\left\{\boxed{7}\right\}
\end{bmatrix}
\right\rangle
\end{bmatrix}
$$

Bare *WH*-Phrases   A bare *wh*-phrase must also have a contextual parameter corresponding to some restricting predicate defined by lexical semantics (in the case of *who*, a restriction that the referent must be a person), and as the overall content is a member of STORE, this

predicate parameter will be the only member of C-PARAMS, as shown in AVM (240).

$$
(240) \quad
\begin{bmatrix}
\textit{wh-phrase} \\
\text{PHON} \quad \left\langle \text{who} \right\rangle \\
\text{CONTENT} \quad \boxed{1}\begin{bmatrix} \textit{wh-param} \\ w : P(w) \end{bmatrix} \\
\text{STORE} \quad \left\{ \boxed{1} \right\} \\
\text{C-PARAMS} \quad \left\{ \begin{bmatrix} P : P = person \end{bmatrix} \right\}
\end{bmatrix}
$$

Notice, however, that this simple specification does not hold to the strong version of the hypothesis as shown. It can easily be held to, though, if we assume that such bare *wh*-phrases have a single syntactic daughter (a bare *wh*-word, as in AVM (241)), and it is therefore the content of this that is being queried by a reprise. Here we take *bare-wh-phrase* as a subtype of *wh-phrase*, and require it to take a daughter with syntactic category *wh-nominal*, a subtype of *nominal*.[74]

$$
(241) \quad
\begin{bmatrix}
\textit{bare-wh-phrase} \\
\text{CONTENT} \quad \boxed{1}\begin{bmatrix} \textit{wh-param} \\ w : P(w) \end{bmatrix} \\
\text{STORE} \quad \left\{ \boxed{1} \right\} \\
\text{C-PARAMS} \quad \boxed{2} \\
\text{HEAD-DTR} \quad \boxed{4}\begin{bmatrix} \text{CAT} \mid \text{HEAD} \quad \textit{wh-nominal} \\ \text{CONTENT} \quad \boxed{5}\begin{bmatrix} P : P = person \end{bmatrix} \\ \text{C-PARAMS} \quad \boxed{2}\left\{ \boxed{5} \right\} \end{bmatrix} \\
\text{DTRS} \quad \left\langle \boxed{4} \right\rangle
\end{bmatrix}
$$

Given the proposed similarity between *wh*-phrases and indefinites, it may be that a similar structure would be suitable for bare indefinites such as *somebody/something*, modulo the CONTENT parameter being of type *non-wh-param*. This certainly seems possible, although it might be that in these cases a branching phrase structure (corresponding to the intuitive morphological decomposition *[some]+[body/thing]*) might allow some further generalisations to be captured.

4.6.3 Verbs and Verb Phrases

Like CNs, verbs are also usually taken to convey predicates – properties of individuals (or of n-tuples of individuals). Should we therefore expect predicate readings for verb reprise questions, as we found with CNs? Are other readings possible (e.g. readings that query an event or situation, or perhaps the arguments of the verb)?

---

[74]Non-*wh* noun phrases must of course have a corresponding restriction to prevent *wh*-words becoming their head nouns.

Given that chapter 3 showed verbs not to be common sources of CRs, we will not go into much detail, but as some verb phrase (VP) sources we observed, and as HPSG sees verbs as the heads of sentences with the content inheritance that this implies, some sort of account is important: this section therefore sketches out an account that seems to be consistent with the data. An initial look at corpus examples does suggest that the predicate readings are available. Event/situation-type readings seem hard to get, and argument readings impossible:

(242)[75]

Unknown:	Have you got any writing paper?
Danny:	No. What for?
Unknown:	Erm I just want to fake a letter.
Danny:	**Fake?**
Unknown:	Yeah.
Danny:	No. `<pause>`

⤳   *"Are you saying it is **faking** you want to do with a letter?"*
⤳   *"What predicate/process do you really mean by 'fake'?"*
⤳   ?#*"When/where/how are you saying you want this faking event to take place?"*
⤳   #*"What are you saying you want to fake?"*

(243)[76]

Joyce:	He had some stuff nicked, a ski jacket which cost me seventy five quid it were half, the rest it should of been a hundred and fifty
Ann:	**Nicked?**
Joyce:	Nicked
Alec:	Mm
Joyce:	Pinched
Ann:	Aargh

⤳   *"Are you saying it was **nicking** that happened to his stuff?"*
⤳   *"What property do you mean by 'nicked'?"*
⤳   ?#*"When/where/how are you saying this nicking event took place?"*
⤳   #*"What are you saying he had nicked?"*

(244)[77]

Martine:	I just heard you, heard you twang– `<laugh>` twanging your ruler. Doing.
Unknown:	**Twanging?** `<pause>`
Martine:	Twanging. `<pause>` Did you, did anybody see that film about erm `<pause>` st– the stolen cars?

⤳   *"Are you saying it is **twanging** you heard?"*
⤳   *"What property do you mean by 'twanging'?"*
⤳   ?#*"When/where/how are you saying I was twanging?"*
⤳   #*"What are you saying I was twanging?"*

---

[75]BNC file KPA, sentences 2684–2690
[76]BNC file KB2, sentences 644–648
[77]BNC file KD8, sentences 3951–3955

As with CNs, then, a representation as predicates does seem reasonable. But as with CNs, this has some interesting consequences for the standard representations of verbs used in HPSG. The usual unification/inheritance-based approach (e.g. G&S) regards verb as denoting *soas*, "states-of-affairs", information not only about the relation between its arguments but about the arguments themselves. In the sentence *"Bo left"*, the content of the verb *left* is the state-of-affairs of Bo leaving, not just the property of leaving. This content is inherited by a VP mother and then the sentence itself.

$$
(245) \quad
\begin{bmatrix}
verb \\
\text{PHON} & \left\langle \text{nicked} \right\rangle \\
\text{CONTENT} &
\begin{bmatrix}
nick\text{-}rel \\
\text{AGENT} & x \\
\text{UNDERGOER} & y \\
\text{EVENT} & e
\end{bmatrix}
\end{bmatrix}
$$

However, if argument readings are impossible, this argues against abstracting this content to C-PARAMS, as the argument indices would go with it. As we saw with CNS, one possible solution might be a representation such as that proposed in (Purver, 2002) and shown in AVM (246), in which the predicate is made available to be added to C-PARAMS and therefore queried:

$$
(246) \quad
\begin{bmatrix}
\text{PHON} & \left\langle \text{nicked} \right\rangle \\
\text{CONTENT} &
\begin{bmatrix}
dyadic\text{-}rel \\
\text{AGENT} & x \\
\text{UNDERGOER} & y \\
\text{PRED} & P
\end{bmatrix} \\
\text{C-PARAMS} & \left\{ \begin{bmatrix} P : name(P, nick) \end{bmatrix} \right\}
\end{bmatrix}
$$

However, this only holds to the weak version of the reprise content hypothesis and therefore doesn't really explain why the argument readings shouldn't be available. A more satisfactory approach might therefore be to take verbs as denoting predicates directly, as in AVM (247). Interestingly, this seems to be moving back towards Montague (1974)'s treatment of verbs as n-place relations ($\lambda x.\lambda y.nick(y, x)$), and also has some similarities to the treatment of verbs in HPSG by the glue-semantic approach of Asudeh and Crouch (2002).

$$
(247) \quad
\begin{bmatrix}
\text{PHON} & \left\langle \text{nicked} \right\rangle \\
\text{CONTENT} & \boxed{1}\begin{bmatrix} P : name(P, nick) \end{bmatrix} \\
\text{C-PARAMS} & \left\{ \boxed{1} \right\}
\end{bmatrix}
$$

Verb Phrases

This must then require VPs to bring the argument structure into the semantic content. For all verbs, a VP would have to combine the verbal predicate with the referent denoted by the sentence subject, which is simple to specify:

$$
(248) \quad
\begin{bmatrix}
\textit{verb-ph} \\[4pt]
\text{CONTENT} & \begin{bmatrix} \textit{soa} \\ \text{AGENT} & \boxed{1} \\ \text{PRED} & \boxed{2} \end{bmatrix} \\[16pt]
\text{HEAD-DTR} & \begin{bmatrix} \textit{verb} \\ \text{CONT} \,|\, \text{INDEX} & \boxed{2} \end{bmatrix} \\[12pt]
\text{SUBJ} & \left\langle \begin{bmatrix} \text{CONT} \,|\, \text{INDEX} & \boxed{1} \end{bmatrix} \right\rangle
\end{bmatrix}
$$

Intransitive, transitive and other types of verbs can then be specified as subtypes of *verb-ph*, and the resulting content can then be inherited by a sentential mother as usual:

$$
(249) \quad
\begin{bmatrix}
\textit{intran-verb-ph} \\[4pt]
\text{CONTENT} & \begin{bmatrix} \textit{monadic-rel} \end{bmatrix} \\[6pt]
\text{COMPS} & \langle \rangle
\end{bmatrix}
\qquad
\begin{bmatrix}
\textit{tran-verb-ph} \\[4pt]
\text{CONTENT} & \begin{bmatrix} \textit{dyadic-rel} \\ \text{UNDERGOER} & \boxed{1} \end{bmatrix} \\[12pt]
\text{COMPS} & \left\langle \begin{bmatrix} \text{CONT} \,|\, \text{INDEX} & \boxed{1} \end{bmatrix} \right\rangle
\end{bmatrix}
$$

The only remaining question is what, if anything, VPs contribute to C-PARAMS. Corpus examples of VP reprises are rare: only a handful were found. Those that were seem to indicate that there is no reading over and above the sub-constituent readings (as suggested in chapter 3): reprises of VPs can be read as querying the verb predicate, or the witness set of any NP arguments, or both, but no other reading seems to be available.

(250)[78]

Ann:	And we had Miss [name] <pause> who used to wear knickers down to her knees.
	She used to sit behind her desk like this with her legs open and her knickers used to come down here, pink ones and blue ones.
	And she used to eat chalk.
Stuart:	**Eat chalk?**
Ann:	Yeah, she was ever so odd. She used to
Stuart:	What, lumps of chalk?
Ann:	she used to chew it.

$\rightsquigarrow$ *"Is it **chalk**$_N$ you're telling me she used to eat N?"*
$\rightsquigarrow$ *"Is it **eating**$_P$ you're telling me she used to do P with chalk?"*
$\rightsquigarrow$ *"Is it **eating**$_P$ and **chalk**$_N$ you're telling me she used to do P with N?"*
$\rightsquigarrow$ *"What do you mean by 'eat chalk'?"*

---

[78]BNC file KB7, sentences 339–346

(251)[79]

	Anon 7:	Where you ever put on the landing?
	Anon 5:	Mhm.
	Anon 6:	Often. `<laugh>`
	Fred:	What for?
	Anon 5:	Hitting people normally.
	Anon 1:	**Hitting people?**
	Anon 5:	Yeah with my pillows. We had pillow fights.

$\rightsquigarrow$ *"Is it **hitting**$_P$ you're telling me you were doing P to people?"*

$\rightsquigarrow$ *"Is it **people**$_N$ you're telling me you were hitting N?"*

$\rightsquigarrow$ *"Is it **hitting**$_P$ and **people**$_N$ you're telling me you were doing P to N?"*

$\rightsquigarrow$ *"What do you mean by 'hitting people'?"*

As can be seen above, the clausal examples seem reasonably clear: either sub-constituent, or both, can be queried, but no other reading seems to be available (there is no way of asking about other arguments such as the subject, for example). It is much less clear what the constituent versions are really asking for, though: if we can paraphrase *"What do you mean by 'eat chalk'?"* as *"What activity and argument are you intending to be conveyed by* "eat chalk"*?"*, or perhaps *"What combination of activity and argument ... "*, then this seems consistent with a sub-constituent analysis – but it's difficult to tell with any certainty.

Given the indications from clausal questions, though, it seems reasonable to propose that VPs amalgamate the C-PARAMS values of their daughters, but contribute nothing else. Whether other elements are contributed to QUANTS (perhaps event variables in a (Davidson, 1980)-like fashion?) is difficult to tell without a larger amount of data, and must therefore be left for future work. This means, of course, that we cannot be sure whether VPs hold to the Definiteness Principle defined for NPs in section 4.4; it seems more conservative to assume not, and instead assume that they just hold to a simple C-PARAMS amalgamation principle, as initially proposed in AVM (191) and repeated here as AVM (252):

$$
(252) \quad
\begin{bmatrix}
phrase \\
\text{C-PARAMS} & \boxed{1} \cup \ldots \cup \boxed{n} \\
\text{DTRS} & \left\langle \begin{bmatrix} \text{C-PARAMS} & \boxed{1} \end{bmatrix}, \ldots, \begin{bmatrix} \text{C-PARAMS} & \boxed{n} \end{bmatrix} \right\rangle
\end{bmatrix}
$$

Indeed, this seems to be applicable to all phrases *except* NPs: no other phrase type seems to allow CRs that query anything other than one or more of their sub-constituents (see the section 4.6.5 below for some prepositional phrase examples). Of course, this seems to be true for sentences too – as with example (3), repeated here as example (253), seems to focus one

---

[79]BNC file KC4, sentences 1310–1317

sub-constituent, in this case a CN:

$(253)^{80}$

A:	I've bought you an aeroplane.
B:	You've bought me an AEROPLANE?
B':	You've bought me a WHAT?

We can therefore take it that the simple inheritance principle of AVM (252) is the default, overridden for NPs by the Definiteness Principle, and that the sub-constituent focussing analysis of section 4.5.4 applies at all levels.

4.6.4 Other Content Words

Chapter 3 suggested that CRs concerning other content words are scarce, with the only other class really worth paying attention to from an empirical point of view being adjectives. The semantic content of an adjective is usually seen as as predicate, either a property of individuals in the case of *intersective* adjectives (*a tall woman*), or possibly a property of other predicates in the case of *non-intersective* adjectives (*the former president*).

All examples found (59 in total) seemed perfectly consistent with a view as predicates: readings concerning e.g. the modified individual were certainly not available.

$(254)^{81}$

George:	I know what when we were dredging down there, we used to have er what we call our safe chains and when we were first dredging down there you'd put your chain in, hands were all purple.
Anon 1:	**Purple?**
George:	All purple.
Anon 1:	Why?
George:	That was the mud, cos they used to put so much sewage into the river `<pause>`
$\leadsto$	*"Is it the property **purple** $_P$ you're telling me your hands had?"*
$\leadsto$	*"Which property are you intending to refer to by 'purple'?"*
$\leadsto$	#*"Which hands are you telling me were all purple?"*

---

[80]From (Blakemore, 1994).
[81]BNC file H5H, sentences 612–616

$(255)^{82}$

> Anon 1: For all that there's no denying the distaste felt by many London lawyers for `<company name>` tactics. They've been seen as too aggressive and too greedy.
>
> Paul: **Aggressive?**
>
> In what sense?
>
> Aggressive in terms that we fight in the market place for for clients and that we er then if that's what aggressive means the answer to that is yes.
>
> $\rightsquigarrow$ *"Is it the property **aggressive** $_P$ you're telling me we're seen as having?"*
> $\rightsquigarrow$ *"Which property exactly are you intending to refer to by 'aggressive'?"*
> $\rightsquigarrow$ #*"Who are you claiming is seen as aggressive?"*

A view as predicates therefore seems sensible. A construction will of course be required to ensure that the semantics of a modified noun get built in a suitable way, along these lines (here, for an intersective adjective – a non-intersective equivalent would presumably involve predicate composition rather than intersection):

$$(256) \quad
\begin{bmatrix}
\text{PHON} & \langle \text{tall, man} \rangle \\
\text{CONT} & \left[P : P = P_1 \cap P_2 \right] \\
\text{HEAD-DTR} & \boxed{1} \\
\text{DTRS} & \left\langle
\begin{bmatrix}
\text{PHON} & \langle \text{tall} \rangle \\
\text{CONT} & \left[P_1 : name(P_1, tall) \right]
\end{bmatrix},
\boxed{1}
\begin{bmatrix}
\text{PHON} & \langle \text{man} \rangle \\
\text{CONT} & \left[P_2 : name(P_2, man) \right]
\end{bmatrix}
\right\rangle
\end{bmatrix}$$

4.6.5 Other Function Words

In chapter 3 it appeared that function words in general (with the exception of number determiners, which have been dealt with in section 4.6.1 above) were very unlikely to form the source of CRs (although they are able to function as *secondary* sources for reprise gaps). The main class we are interested in providing an analysis for is prepositions: some CR sources were seen to be prepositional phrases (PPs), although these seemed to be focussing on NP sub-constituents. Conjunctions and complementisers are not common enough as CR sources to be important for the current purposes.

Searching for reprises of prepositions in the BNC confirms this: of 8 examples found, 5 were reprise gaps (see example (257)), 2 were reprises of original utterance-referring or

---

[82]BNC file HMJ, sentences 89–93

utterance-anaphoric sources rather than standard prepositions (see example (258)):

(257)[83]

Unknown:	What exactly do you mean by unpreparedness?
Ken:	**By?**
Unknown:	Unpreparedness.
Ken:	Well, the fact that \<pause> in a general election situation, people will go out and vote. In local elections they won't.

$\leadsto$ *"What word did you say after 'by'?"*

(258)[84]

Jean:	Can anybody spell \<pause> er \<pause> between?
Unknown:	**Between?**
Jean:	Yes Candice? Stand up and see if you can spell between.
Unknown:	\<spelling phonetically> B E T W E E N

$\leadsto$ *"Is it the word 'between' you're asking if anybody can spell?"*

Only one example seemed ambiguous between being a reprise gap and a clausal reprise fragment (example (259)), and in this case a view of prepositions as denoting logical relations, as with determiners, seems reasonable:

(259)[85]

Dave:	I don't know any of the \<pause> except on there, and that, that were \<pause> and tha- that's, that's before er October.
Margaret:	**Before?**
Dave:	That's way before October.
Margaret:	Well that's last year's!

$\leadsto$ *"What word did you say after 'before'?"*
$\leadsto$ *"Is it the relation **before**$_R$ you're telling me it's $R(October)$?"*

A similar analysis to determiners is therefore proposed, whereby function words do contribute their content, a logical relation, to C-PARAMS, but as this content is very easily identifiable and commonly known, they are extremely unlikely to cause CRs.

Prepositional phrases can now be analysed along the same lines as VPs and other content phrases: they contribute nothing to C-PARAMS beyond the parameters already contributed by their daughters. In this case, this would mean they have parameters associated with the preposition and the NP, and as prepositions are unlikely to cause CRs, we should find that PP reprises all query their NP daughters. There aren't many examples, but this does indeed seem

---

[83]BNC file F7T, sentences 359–363
[84]BNC file KCK, sentences 655–659
[85]BNC file KD2, sentences 1758–1761

to be the case in all of them:

$(260)^{86}$

A.:	. . . I tell you what there's another thing on television that makes me re-ally laugh is that advert about electricity and it says
Arthur:	**About electricity?**
A.:	Electricity yes
Arthur:	Yeah

$\rightsquigarrow$ *"Is it **electricity** you're telling me there's that advert about?"*

$(261)^{87}$

SB:	And going ever smaller, can you get from neutrinors to smaller parti-cles?
TN:	**From neutrinors?**
	Well we haven't really discussed neutrionors yet. We should leave the neutrionors aside for the time being.

$\rightsquigarrow$ *"Is it **neutrinos** you're asking if we can get from?"*

4.7 Conclusions

This chapter has introduced the use of reprise questions as probes for investigating the semantic content of words and phrases, giving a strong criterion of assigning denotations which not only combine to make up compositional sentence meanings but explain why individual constituents give their observed reprise readings. We have examined the evidence provided by the apparent meaning of these questions as regards the semantic content of nouns and noun phrases, and (very briefly) other content and function words. This evidence has led to the following conclusions:

- The commonly held view of CNs as properties (of individuals) seems to correspond well with their reprises. This seems to hold for most content words.

- The view of NPs as denoting sets of sets, or properties of properties, seem very difficult to reconcile with reprise questions.

- Reprises of all phrases seem to be able to query focussed sub-constituents. NPs seem to be able to query something else too – their own content.

- Reprises of definite NPs suggest that most uses of these NPs are referential to a (possibly functional) individual or set.

- Reprises of indefinite NPs and other QNPs suggest that such referential uses, while rare, are possible.

---

[86]BNC file KP1, sentences 1367–1370
[87]BNC file KRH, sentences 3024–3027

- Reprises of function words are rare, but those that exist seem to query logical relations.

These conclusions strongly suggest a representation of NPs as denoting witness sets, and a definite/indefinite distinction expressed by abstraction of these sets to C-PARAMS (or lack thereof). Brief accounts have been given of relative quantifier scope via a functional view, intrasentential anaphora via a parameter binding mechanism, non-monotone-increasing quantifiers via a representation as pairs of sets, and sub-constituent focussing via a link between information structure and MAX-QUD.

We have also seen along the way that these conclusions cause us to revise some of the standard assumptions made in HPSG about inheritance of content and other features (like C-PARAMS) from daughters to mothers.

Chapter 5

A Grammar for Clarification

5.1 Introduction

We now have an ontology of the possible forms and readings of CRs, together with a reasonable idea of how a suitable grammar must behave with respect to semantic inheritance and contextual dependence. This chapter now puts together the basic insights of the G&C analysis of clarification, together with the ontology developed in chapter 3 and the semantic framework of chapter 4 into a HPSG grammar fragment which gives an analysis of the major clarificational forms and readings that we are concerned with, and which can then be used in the CLARIE dialogue system described in chapter 6. It also incorporates, and modifies, the basic approach to elliptical fragments (and their reconstruction) of the SHARDS system described in chapter 2.

Firstly, section 5.2 describes some of the main distinguishing features of the grammar that are required for CR analysis. Section 5.3 then outlines a treatment of elliptical fragments that is consistent with the approach and observations so far. Section 5.4 then shows how these features are combined to treat the CR forms and readings we have seen.

Coverage Note that the intention of this grammar is to provide an analysis for the various types of CR (together with a general approach that explains the clarificatory potential of source utterances in general). It has been implemented and is used for interpretation and generation by the CLARIE system, but the purpose of that system is similarly to demonstrate clarificational capabilities rather than broad coverage. Given this, there will be no discussion here of coverage of other linguistic phenomena and constructions, and indeed the general coverage of the grammar as currently implemented is poor, although there seems no reason to believe that broad coverage could not be achieved given a suitable set of lexical and phrasal type definitions.

5.2   Basic Requirements

As will have become clear by now, there are several basic requirements that a grammar must meet in order to be suitable for analysing CRs. Firstly, it must include information at phonological, syntactic and semantic levels – this is met by the use of HPSG here, although it is quite possible that other frameworks could be used (for example, see Cooper and Ginzburg, 2002, for a formulation of some of G&C's account in Martin-Löf type theory). Other requirements that have already been mentioned are contextual abstraction, the representation of conversational move type, and utterance anaphora: the incorporation of these into the grammar is described here in section 5.2.1, section 5.2.2 and section 5.2.3 respectively.

While we have not paid it much attention so far, a further requirement is that the grammar must be able to handle unknown words – as mentioned in chapter 2, and as will be discussed in detail in chapter 6, one of the motivations behind a treatment of CRs is to allow a dialogue system to cope with and ask about unknown words. This is described in section 5.2.4.

Of course, one other major requirement is the ability to handle elliptical fragments (as CRs are so often elliptical) – this is left for section 5.3 where it is discussed in detail.

5.2.1   Contextual Abstraction

The first major point of departure from a standard HPSG grammar is of course the use of the C-PARAMS feature to express contextual abstraction, while incorporating the functions of the C-INDICES and/or BACKGROUND features that might be found in other more standard grammars. It is the identification of the members of C-PARAMS, or rather problems therewith, that give rise to clarification.

The grammar assumes that all standard words contribute their semantic content to C-PARAMS as a contextual parameter. In the case of PNs, the referent of this parameter (its INDEX value) is the individual who bears the associated name. As outlined in chapter 4, the referent of parameters associated with CNs and verbs is taken to be a predicate, which again bears a particular name. Adjectives and adverbs are assumed to follow the same principle and refer to named predicates. Function words such as prepositions and determiners denote logical relations. This general principle can be expressed as a simple constraint in the grammar on signs of type *lex* (i.e. on words, as opposed to phrases):

$$(262) \quad \begin{bmatrix} lex \\ \text{CONT} & \boxed{1}\begin{bmatrix} parameter \end{bmatrix} \\ \text{C-PARAMS} & \left\{ \boxed{1} \right\} \end{bmatrix}$$

The only exceptions to this constraint are words such as greetings and conventional CRs, which are associated directly with sentential content (dialogue moves) in the lexicon, rather than having a parameter as content – see below. Typical entries for a PN and CN are shown

in (263):

$$
(263)\quad
\begin{bmatrix}
\textit{lex \& name} \\
\text{PHON} \quad \langle \text{paris} \rangle \\
\text{CONT} \quad \boxed{1}
\begin{bmatrix}
\textit{parameter} \\
\text{INDEX} \quad \boxed{1}\textit{object} \\
\text{RESTR} \quad \left\{ \begin{bmatrix} \text{INSTANCE} \quad \boxed{1} \\ \text{NAME} \quad \text{paris} \end{bmatrix} \right\}
\end{bmatrix} \\
\text{C-PARAMS} \quad \left\{ \boxed{1} \right\}
\end{bmatrix}
\quad
\begin{bmatrix}
\textit{lex \& noun} \\
\text{PHON} \quad \langle \text{ticket} \rangle \\
\text{CONT} \quad \boxed{1}
\begin{bmatrix}
\textit{parameter} \\
\text{INDEX} \quad \boxed{1}\textit{predicate} \\
\text{RESTR} \quad \left\{ \begin{bmatrix} \text{INSTANCE} \quad \boxed{1} \\ \text{NAME} \quad \text{ticket} \end{bmatrix} \right\}
\end{bmatrix} \\
\text{C-PARAMS} \quad \left\{ \boxed{1} \right\}
\end{bmatrix}
$$
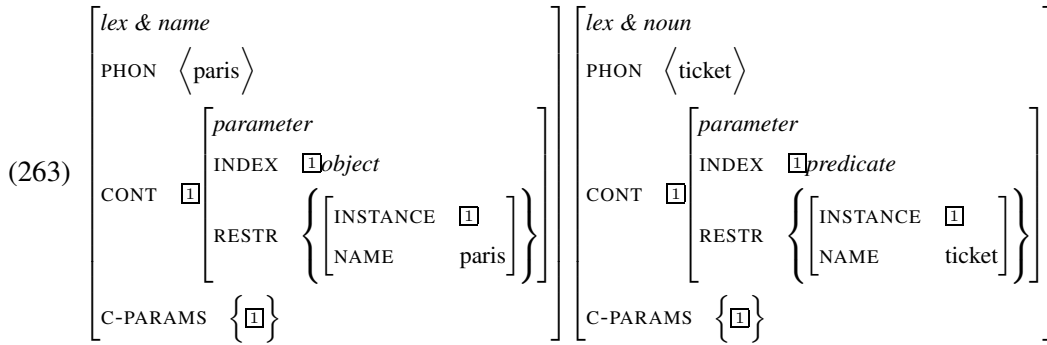
The scarcity of function word CRs (see chapter 3) suggests an alternative approach whereby function words are *not* given associated contextual parameters. If a system/grammar is not intended to treat function words as being able to give rise to CRs, their content would be taken as given, rather than having to be identified in context. This would prevent an analysis of clarification of function words, but might make for a simpler grammar. However, there are reasons not to take this approach. While chapter 3 has shown that function word CRs are rare, it has not shown that they are impossible; indeed, chapter 4 showed that for some function words (e.g. numerical determiners), CRs (at least with clausal readings) seem to be quite natural. If a grammar excludes the possibility of function word CRs, it will have to be re-worked for any system that later wants to be able to take them into account.

A more modular approach which would involve less re-work is instead to allow the grammar to produce analyses consistent with function word CRs, and then use a separate module to exclude or allow them. A change in overall strategy will then only involve re-working this module rather than the whole grammar. This approach will be taken here and in the dialogue system of chapter 6, where this separate module will be that part of the dialogue move engine (DME) which defines the grounding process – see section 6.3. This will have two implications for the grammar: function words will be made contextually dependent, projecting C-PARAMS and thereby allowing them to be sources of CRs in theory (although the grounding module will make their parameters easy to ground); and function word fragments will be given analyses as CRs themselves (although the grounding module will prefer other analyses).

As described in chapter 4, the contextual parameters are amalgamated across syntactic daughters, allowing them all to percolate up to the top level of utterance representation. On the way, parameters are added by definite NPs, following the analysis given in chapter 4. At the top level, parameters for speaker and addressee are also added (these are required both to express the full contextual dependence of the utterance, and to give its full content – see section 5.2.2 below). A CONSTITS feature is also amalgamated (see Ginzburg and Cooper, 2004) so that information about which constituent contributed which contextual parameter is

available. The amalgamated C-PARAMS value of a sentence then looks as follows:

$$(264) \quad \begin{bmatrix} \text{PHON} & \langle \text{john, likes, mary} \rangle \\ \text{CONT} & \left[assert(i, j, P(x, y)) \right] \\ \text{C-PARAMS} & \left\{ \begin{array}{l} \boxed{1}\left[x\!:\!name(x, john) \right], \boxed{2}\left[P\!:\!name(P, like) \right], \\ \boxed{3}\left[y\!:\!name(y, mary) \right], \\ \left[i\!:\!spkr(i) \right], \left[j\!:\!addr(j) \right] \end{array} \right\} \\ \text{CONSTITS} & \left\{ \begin{bmatrix} \text{PHON} & \langle \text{john} \rangle \\ \text{CONT} & \boxed{1} \end{bmatrix}, \begin{bmatrix} \text{PHON} & \langle \text{likes} \rangle \\ \text{CONT} & \boxed{2} \end{bmatrix}, \begin{bmatrix} \text{PHON} & \langle \text{mary} \rangle \\ \text{CONT} & \boxed{3} \end{bmatrix} \right\} \end{bmatrix}$$

This provides the representation we need to account for all the possible sources of clarification in an utterance.

5.2.2   Conversational Move Type

A second point of departure is the inclusion of move type. A sign produced by the basic SHARDS grammar (and by most typical HPSG grammars) has as its content the logical proposition (or question etc.) associated with the utterance – see AVM (265).

$$(265) \quad \begin{bmatrix} interrogative\text{-}cl \\ \text{PHON} & \langle \text{does, john, like, mary} \rangle \\ \text{CONT} & \left[?.like(x, y) \right] \\ \text{BACKGROUND} & \left\{ \left[x\!:\!name(x, john) \right], \left[y\!:\!name(y, mary) \right] \right\} \end{bmatrix}$$

In contrast, G&C's analysis (following G&S) assumes that the grammar assigns conversational move type, so that the semantic content of an utterance includes its basic illocutionary force: rather than the question *whether John likes Mary* (as shown above), the content becomes the proposition that *A is asking B whether John likes Mary*. This is required in order to account for the way that clausal CRs derive their content - they are querying the move made by the previous utterance. In a dialogue system, it could be argued to have another advantage: if the content of an utterance is a move, it can be passed directly to the DME for use in updating the information state (IS), as typical IS update rules depend on move type as well as propositional content (whether the last move made is e.g. asking a question, asserting an answer, greeting the user etc.).[1]

A grammatical treatment has already been given by (Ginzburg et al., 2001b, 2003) and can be followed directly: top-level root clauses (specified as being of type *root-cl(ause)*)

---

[1]Another motivation behind such an approach is given by (Ginzburg et al., 2001b, 2003) as the ability to analyse conventional phrases such as greetings, which have little or no propositional content and can only be represented as conversational moves – see below. Another might be the existence of certain constructions such as the bare *"Why?"* which appear to refer to an antecedent move as in example (266), although see (Ginzburg,

are defined to include move type as part of their propositional content, with this move type determined from syntactic/semantic form of their head daughter (declarative sentences are treated as moves of type *assert*, interrogatives as type *ask*, etc. as shown in AVM (269)).

$$(269)\quad \begin{bmatrix} \textit{root-cl} \\ \text{PHON} \quad \left\langle \text{does, john, like, mary} \right\rangle \\ \text{CONT} \quad \left[ask(i, j, \boxed{1}\left[?.P(x,y)\right]) \right] \\ \text{HEAD-DTR} \quad \begin{bmatrix} \textit{interrogative-cl} \\ \text{CONT} \quad \boxed{1} \end{bmatrix} \\ \text{C-PARAMS} \quad \left\{ \begin{array}{l} \left[x\!:\!name(x, john)\right], \left[P\!:\!name(P, like)\right], \\ \left[y\!:\!name(y, mary)\right], \left[i\!:\!spkr(i)\right], \left[j\!:\!addr(j)\right] \end{array} \right\} \end{bmatrix}$$

Specifically, this is achieved for standard sentences by constraining the type *root-cl* to have a move as content (an object of type *illoc-rel*) and to take the content of the head daughter as the message argument of that move (the value of its MSG-ARG feature – see AVM (270)). As already mentioned, root clauses also introduce C-PARAMS members relating to the identity

2003) for more discussion of this.

(266) | A: Did Bo leave?
       | B: Why?
       | ↝   *"Why are you asking me whether Bo left?"*

While the apparent unavailability of reference to the move in example (267) might be argued to provide counter-evidence, as example (268) shows, this reference is in fact available given a suitable example.

(267) | A: John's going shopping.
       | B: That's surprising.
       | ↝   *"It's surprising that John's going shopping."*
       | ↝   #*"It's surprising that you're asserting that John's going shopping."*

(268) | A: You've got a big nose.
       | B: That's very rude.
       | ↝   #*"It's rude that I've got a big nose."*
       | ↝   *"It's rude that you're asserting that I've got a big nose."*

of speaker and addressee.[2]

$$
(270) \quad
\begin{bmatrix}
\textit{root-cl} \\[2pt]
\text{CONT} \quad
\begin{bmatrix}
\textit{proposition} \\
\text{QUANTS} \quad \{\} \\[4pt]
\text{NUCL} \quad
\begin{bmatrix}
\textit{illoc-rel} \\
\text{UTT} \quad \boxed{i} \\
\text{ADD} \quad \boxed{j} \\
\text{MSG-ARG} \quad \boxed{1}
\end{bmatrix}
\end{bmatrix} \\[4pt]
\text{HEAD-DTR} \mid \text{CONT} \quad \boxed{1} \\[2pt]
\text{C-PARAMS} \quad \left\{ \left[\boxed{i} : spkr(\boxed{i})\right], \left[\boxed{j} : addr(\boxed{j})\right] \right\}
\end{bmatrix}
$$
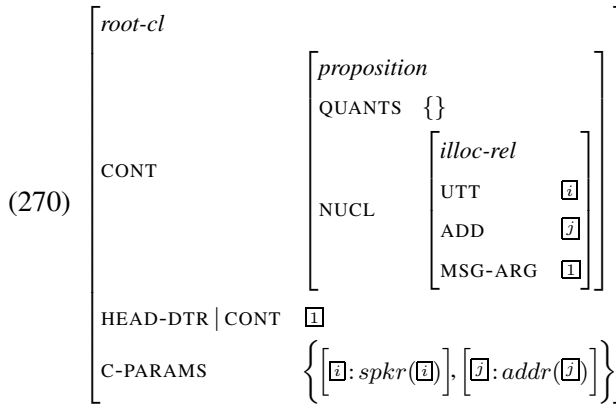
Particular move types (subtypes of *illoc-rel*) are associated with particular message types, thus forcing a clause whose head daughter content is a question to be an *ask* move, one whose head daughter is a proposition an *assert* move, and so on:

$$
(271) \quad
\begin{bmatrix}
\textit{assert-rel} \\
\text{MSG-ARG} \quad \left[\textit{proposition}\right]
\end{bmatrix}
\qquad
\begin{bmatrix}
\textit{ask-rel} \\
\text{MSG-ARG} \quad \left[\textit{question}\right]
\end{bmatrix}
$$

For conventional words or phrases which have only a move as their content (such as greetings, interjections etc., which are not usually regarded as having standard propositional content) the grammar again follows (Ginzburg et al., 2003), and the move type is associated directly with content in the lexicon (see AVM (272)).

$$
(272) \quad
\begin{bmatrix}
\text{PHON} \quad \left\langle \text{hello} \right\rangle \\[4pt]
\text{CONT} \quad
\begin{bmatrix}
\textit{greet-rel} \\
\text{SPKR} \quad \boxed{1} \\
\text{ADDR} \quad \boxed{2}
\end{bmatrix} \\[4pt]
\text{C-PARAMS} \quad \left\{ \left[\boxed{1} : spkr(\boxed{1})\right], \left[\boxed{2} : addr(\boxed{2})\right] \right\}
\end{bmatrix}
$$
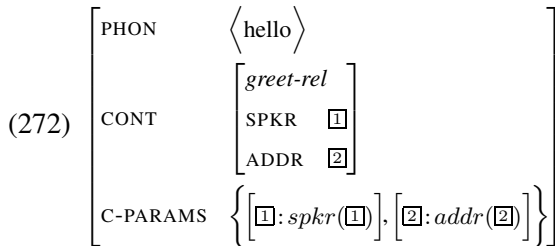
Note that moves are only divided at this stage into a simple hierarchy as shown in figure 5.1: four types that can be distinguished by message type, and a hierarchy of those that are specified conventionally in the lexicon. Most dialogue systems distinguish move type to a finer grain than this, distinguishing for example *assertions* from *answers*, which might be further sub-divided into, say, *positive answers* and *negative answers*, as these will have distinct IS update effects. As will be described in chapter 6, this distinction will be made at the stage of grounding and IS update, when general IS context can be taken into account, rather than assuming that it is specified by the grammar or in the lexicon.

---

[2]Other contextual details such as utterance time and location might also be given parameters at this point, but are left out here for clarity's sake.
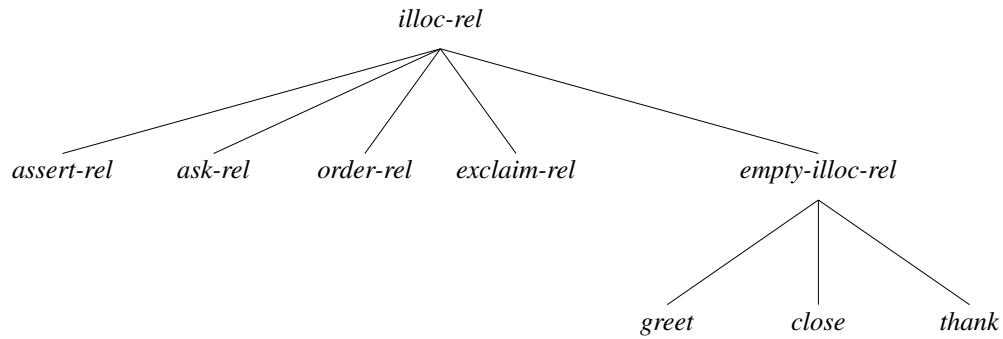
Figure 5.1: Conversational Move Type Hierarchy

5.2.3   Utterance Anaphora

Again following G&C, the grammar must also include the *utt(erance)-anaph(oric)-ph(rase)* type required by both constituent and lexical readings. The utterance-anaphoric phrase type is therefore defined as follows, to denote any contextually salient utterance which has the same phonological form:

(273)
$$
\begin{bmatrix}
\textit{utt-anaph-ph} \\
\text{PHON} \quad\quad\quad\quad \boxed{1} \\
\text{CONT} \quad\quad\quad\quad \boxed{2} \\
\text{CTXT} \,|\, \text{SAL-UTT} \quad \boxed{2}\begin{bmatrix}\text{PHON} \quad \boxed{1}\end{bmatrix} \\
\text{HEAD-DTR} \quad\quad \begin{bmatrix}\textit{lex}\end{bmatrix}
\end{bmatrix}
$$

Note that the definition above allows any word (any sign of type *lex*) to be taken as the head daughter. Although G&C's account of utterance anaphora is only defined for NPs (and only explicitly laid out for single-word PNs), it can in principle be applied to any phrase type, and this does seem to be necessary: chapter 4 has shown examples of constituent reprise questions for multi-word NPs and for content phrases other than NPs, although not for function words. However, for lexical readings and the gap and filler forms, it seems likely that any arbitrary substring (not necessarily a standard constituent of the grammar) can be used utterance-anaphorically, so the grammar must allow this.

This means that for any sentence, the grammar must be able to construct utterance-anaphoric phrases corresponding to all $n(n+1)/2$ possible substrings, from individual words up to the entire sentence. This can be achieved by allowing multi-word utterance-anaphoric phrases to be built from smaller ones. First, the type *utt-anaph-ph* is split into two subtypes: *lex-utt-anaph-ph*, which forms an utterance-anaphoric phrase from any single word; and *phr-utt-anaph-ph*, which then allows multi-word versions to be built.

The single-word type, shown in AVM (274), allows an utterance-anaphoric phrase to be built directly from the head daughter word, essentially following G&C's original specifica-

tion, except that of course the daughter can now be any single word (rather than just any NP). The constraints inherited from the *utt-anaph-ph* supertype are shown shaded:

$$(274) \quad \begin{bmatrix} \textit{lex-utt-anaph-ph} \\ \text{PHON} & \boxed{1} \\ \text{CONT} & \boxed{2} \\ \text{CTXT} \mid \text{SAL-UTT} & \boxed{2}\begin{bmatrix} \text{PHON} & \boxed{1} \end{bmatrix} \\ \text{HEAD-DTR} & \boxed{3}\begin{bmatrix} \textit{lex} \\ \text{PHON} & \boxed{1} \end{bmatrix} \\ \text{DTRS} & \left\langle \boxed{3} \right\rangle \end{bmatrix}$$
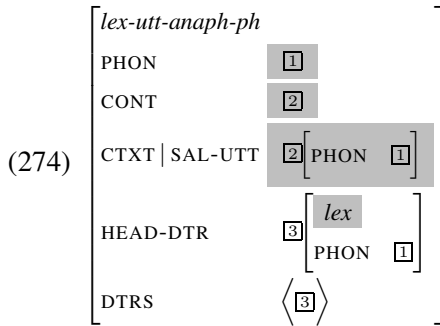
The multi-word type, shown in AVM (275), allows an utterance-anaphoric phrase to be built from any word combined with an existing utterance-anaphoric phrase.

$$(275) \quad \begin{bmatrix} \textit{phr-utt-anaph-ph} \\ \text{PHON} & \boxed{1} \; \left\langle \boxed{4} \oplus \boxed{5} \right\rangle \\ \text{CONT} & \boxed{2} \\ \text{CTXT} \mid \text{SAL-UTT} & \boxed{2}\begin{bmatrix} \text{PHON} & \boxed{1} \end{bmatrix} \\ \text{HEAD-DTR} & \boxed{3}\begin{bmatrix} \textit{lex} \\ \text{PHON} & \boxed{4} \end{bmatrix} \\ \text{DTRS} & \left\langle \boxed{3}, \begin{bmatrix} \textit{utt-anaph-ph} \\ \text{PHON} & \boxed{5} \end{bmatrix} \right\rangle \end{bmatrix}$$
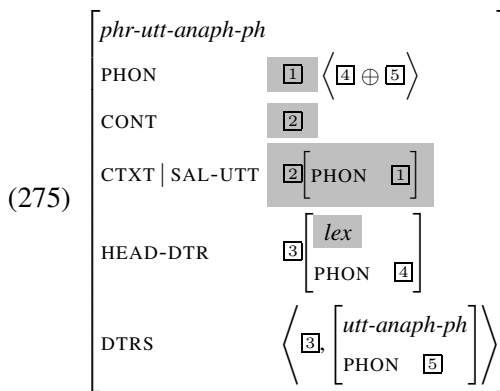
Taken together, the two types will allow an utterance-anaphoric phrase to be built for any substring, which will always denote a phonologically parallel SAL-UTT. It might be possible to argue against this approach on the grounds that it introduces a construction that is non-compositional: the content of the mother is not a combination of the content of its daughters, but a separate sign altogether. But this seems unavoidable given the nature of utterance-anaphora; it's not clear whether compositionality should really apply to utterance-anaphoric phrases in the first place, of course – and the same argument applies to G&C's original approach.[3]

The result of this addition is that the grammar can now assign two alternative representations to a typical sentence, one with the standard semantics, and one anaphorically referring to an utterance with the same surface word string:

---

[3]One alternative way to build these phrases might be to allow the parser to directly create utterance-anaphoric edges corresponding to all possible substrings, thus avoiding the need to build one utterance-anaphoric phrase from another. However, this seems arbitrary and would remove the independence of grammar from parser.

$$(276) \quad \begin{bmatrix} \textit{interrogative-cl} \\ \text{PHON} \quad \left\langle \text{does, john, like, mary} \right\rangle \\ \text{CONT} \quad ?.P(x,y) \\ \text{CTXT} \,|\, \text{C-PARAMS} \quad \left\{ \begin{bmatrix} x : name(x, john) \end{bmatrix}, \begin{bmatrix} P : name(P, like) \end{bmatrix}, \begin{bmatrix} y : name(y, mary) \end{bmatrix} \right\} \end{bmatrix} \begin{bmatrix} \textit{utt-anaph-ph} \\ \text{PHON} \quad \boxed{1} \left\langle \text{does, john, like, mary} \right\rangle \\ \text{CONT} \quad \boxed{2} \\ \text{CTXT} \,|\, \text{SAL-UTT} \quad \boxed{2} \begin{bmatrix} \text{PHON} \quad \boxed{1} \end{bmatrix} \end{bmatrix}$$

Note that the utterance-anaphoric version does not contribute to C-PARAMS in the same way as the standard version. This is the desired behaviour: the content of the anaphoric version is purely the previously uttered sign, and so does not involve the reference to named individuals (amongst other things) that the standard version does. The reference to a previous utterance *"does John like Mary?"* does not involve establishing the referents of *John* or *Mary*. It does, however, require establishing which utterance in context is being referred to, so might be expected to project a contextual parameter corresponding to that utterance which must be grounded: this is discussed further in section 5.3 where exactly this analysis will be proposed. Note here, though, that this requires utterance-anaphoric phrases to override the default C-PARAMS amalgamation principle, as those parameters associated with the lexical words (e.g. *john*) are not inherited.

5.2.4   Unknown Words

If a dialogue system is to handle sentences containing unknown words, a necessary first step must be for the grammar to be able to produce a syntactic and semantic analysis of such sentences, which can be passed on to other modules for grounding (which will presumably fail) and subsequent clarification. The representation produced by the grammar must be as full as possible, with syntactic structure and semantic argument structure specified, so that when ensuing clarificational dialogue has established the meaning of the unknown word, the original sentence receives a full interpretation.

Syntax

Parsing can be achieved by allowing words not in the lexicon to be represented as a disjunction of generic entries for all open-class syntactic categories, following Erbach (1990). For example, entries for CN, intransitive verb, transitive verb and so on will be included, but not entries for closed classes (unknown prepositions and determiners can be assumed not to exist). The entries are generic in that only basic syntactic selectional restrictions and semantic predicate-argument structure are included: the more detailed selectional constraints that are common in lexicalized grammars and that stem from lexical semantics (e.g. restriction of particular argument positions to specific semantic classes) cannot of course be specified. Ex-

amination of the final state of the parser will determine possible correct syntactic categories: essentially those that produce the set of successful parses.

At this stage, syntactic ambiguity is likely (for example, in the sentence *"I saw her X"*, X could be a noun or a verb). This ambiguity could be reduced by part-of-speech (PoS) tagging based on orthographic form prior to parsing (there are many common suffixes which indicate nouns or verbs, and PoS-taggers exploit these very successfully (see e.g. Ratnaparkhi, 1996)). Although this has not been implemented in the current system as the grammar is small enough for ambiguity not to be a serious concern, it would almost certainly be required with a larger grammar (but should not be difficult to implement). Some ambiguity is unavoidable: the set of alternative parses is kept throughout processing of the utterance, until clarification of the unknown word resolves it.

Semantics

The semantic analysis of nouns and verbs set out in chapter 4 has the nice additional property of straightforwardly giving a suitable representation for unknown words. The content of a normal CN is taken to be a parameter referential to a named predicate: that of an unknown word assumed to be a CN is exactly the same, with the name taken directly from the orthography of the word.

$$
(277) \quad
\begin{bmatrix}
\text{PHON} & \left\langle \text{heffalump} \right\rangle \\[4pt]
\text{CONTENT} & \boxed{1} \\[10pt]
\text{C-PARAMS} &
\left\{ \boxed{1}
\begin{bmatrix}
\textit{parameter} \\
\text{INDEX} & P \\[6pt]
\text{RESTR} &
\left\{
\begin{bmatrix}
\text{INSTANCE} & P \\
\text{NAME} & \text{heffalump}
\end{bmatrix}
\right\}
\end{bmatrix}
\right\}
\end{bmatrix}
$$

For inflected forms (e.g. plural nouns, third person singular or past tense verbs), simple morphological rules can be used for stemming: to produce the root form of the word for use as the predicate name. While some words inflect irregularly, in English at least they are not only relatively few in number (an examination of the OALD (Hornby, 1974) shows that a simple set of 16 plural-formation rules account for over 99% of nouns, and 6 rules account for over 99% of third person singular verb forms) but are also common in usage (and thus unlikely to be outside the system lexicon to begin with).

Note that this representation, while it may leave out unknown information and/or restrictions such as animacy, is not in itself underspecified – no features are left without values. There is therefore no *semantic* distinction between an unknown word and a known one:[4] the difference will come in the grounding process, when referents for unknown predicates

---

[4]Syntactically, of course, there will be: we do not know the category of the unknown word, so entries for noun, verb, adjective etc. will be created.

will not be found. The overall semantic content of the sentence, and its predicate-argument structure, can therefore be built in full, avoiding much of the ambiguity of semantic structure determination described by Knight (1996).

5.3   Elliptical Fragments

The original SHARDS grammar produces a sign representation which, in the case of elliptical fragments, is underspecified. This is passed to the ellipsis reconstruction module, which uses contextual information (a set of possible questions under discussion (QUDs) and a set of possible salient utterances (SAL-UTTs) calculated using a set of simple dialogue processing rules) to fully specify the sign.

5.3.1   Contextual Specification

As (Schlangen, 2003; Schlangen and Lascarides, 2003) point out, this modular method involves underspecification of the contextual features MAX-QUD and SAL-UTT during parsing: the parser produces a sign which leaves these values undefined, and they are then determined by the separate ellipsis reconstruction module according to the possible values provided by context. This goes against the standard view of HPSG parsing, which requires all features to be *ground* i.e. fully specified in the output sign, and therefore prevents use of the standard (and robust and efficient) parsers that have been developed for large-scale grammars, e.g. the English Resource Grammar (Copestake and Flickinger, 2000).

The alternative would be to assume that MAX-QUD and SAL-UTT are given before parsing, so that their values can be fully specified during parsing, and a standard parser could be used. Schlangen regards this as untenable: possible values of both features may depend on the results of parsing, and reasoning about the parsed sign in context, and cannot therefore be provided before parsing begins (unless all possible values are produced, potentially a very large number). Particularly when clarification is considered, we can see that this approach would be unintuitive: values of MAX-QUD stemming from coercion operations on context which relate to particular signs would have to be produced before those signs have been generated by the parser. It could also not be an efficient approach: possible coercion operations (and resulting MAX-QUD and SAL-UTT values) corresponding to all substrings of any string would have to be considered. Schlangen also regards other issues as problematic, perhaps most importantly that this approach is not strictly compositional (in that the content of a fragment does not stem purely from its constituent words but from context) and that it no longer regards the CONTEXT feature as providing a restriction on the use of a sign as originally intended (Pollard and Sag, 1994), but as a direct input to the sign's semantic content.

Resolution of ellipsis *after* parsing therefore seems preferable, when possible contextual feature values can be considered based on both the sign itself and the context. The process of

ellipsis resolution then consists of finding appropriate possible values for the MAX-QUD and SAL-UTT features in context which can produce a fully-specified sign which is both internally and contextually consistent. But this seems directly parallel to the definition of the grounding process for C-PARAMS: grounding an utterance abstract consists of finding appropriate possible values for the members of C-PARAMS which produce a fully-specified consistent sign.

So a natural move therefore seems to be to consider MAX-QUD and SAL-UTT as members of C-PARAMS, as shown in AVM (278) for a short answer. Given that C-PARAMS is intended to express an utterance's contextual dependence (it is the abstracted set in the representation of an utterance as a contextually-dependent abstract), this seems suitable: elliptical utterances are directly dependent on these two features of context. It would also offer a solution to the specification problem described above: the values of these features are no longer simply left underspecified, but form part of an abstracted set. The sign can therefore be fully ground, and as such produced by standard parsers, without having to specify the values of the features.

$$
(278) \quad
\begin{bmatrix}
\text{PHON} & \langle \text{john} \rangle \\[4pt]
\text{CONT} & \boxed{1}\big[\textit{proposition}\big] \\[4pt]
\text{HEAD-DTR} & \Big[\text{CONT} \quad \big[\boxed{3}:name(\boxed{3},john)\big]\Big] \\[6pt]
\text{CTXT} &
\begin{bmatrix}
\text{MAX-QUD} & \boxed{4}\big[\text{PROP} \quad \boxed{1}\big] \\[4pt]
\text{SAL-UTT} & \boxed{5}\big[\text{CONT}\,|\,\text{INDEX} \quad \boxed{3}\big]
\end{bmatrix} \\[14pt]
\text{C-PARAMS} &
\left\{
\ldots,
\begin{bmatrix}
\textit{parameter} \\
\text{INDEX} \quad \boxed{4}\big[\textit{question}\big] \\
\text{RESTR} \quad \big\{max\_qud(\boxed{4})\big\}
\end{bmatrix},
\begin{bmatrix}
\textit{parameter} \\
\text{INDEX} \quad \boxed{5}\big[\textit{sign}\big] \\
\text{RESTR} \quad \big\{sal\_utt(\boxed{5})\big\}
\end{bmatrix},
\ldots
\right\}
\end{bmatrix}
$$

It is clear that this move requires the INDEX feature of a parameter to be typed such that objects such as questions and signs can be suitable values, but note that this move has already been made by G&C's analysis of utterance-anaphora (where a sign is taken to be the referential content of a parameter). Note also that the CONTEXT feature cannot be done away with: it still plays the role of expressing constraints between various features of the sign that depend on features of context, it is just that its features are now abstracted to C-PARAMS so that they can be instantiated with contextually provided values during grounding.

This abstraction approach[5] appears not to suffer from the problems outlined above. Firstly, it allows the representation of a fragment to be entirely compositional; the semantic content is an abstract, of which the non-abstracted parts are derived entirely from the constituent words. Secondly, the representation is contextually dependent but leaves nothing underspecified so

---

[5]Note that the use of abstraction here is not the same as the higher-order abstraction approach of (Dalrymple et al., 1991) used for VP ellipsis, in which abstracts are formed from the antecedent utterance and used in resolving the elliptical sentence. Here, the elliptical fragment is seen as the abstract, to be applied to the context.

is consistent with standard parsing routines and a standard parser can be used (provided that it can deal with abstracts as objects, but this capability is already required for other reasons – for one thing, questions are represented as abstracts, not only here but in the English Resource Grammar). Thirdly, the role of the CONTEXT feature is once again to express constraints on the use of the sign (rather than supply external information during parsing): it describes the kind of context to which the abstract can be successfully applied. Finally, the reconstruction of the full sign in context occurs after parsing during the grounding process, and therefore properties of both sign and context, and reasoning with and about them, can be used without re-parsing being required.

At first sight this approach might appear to have the disadvantage of spreading the process of ellipsis resolution between both the linguistic module (the grammar) and the DME (grounding), something that has been argued against by e.g. Lewin and Pulman (1995) as it can reduce the modularity of an implementation. However, this is not the case here: modularity is retained, as the two processes are separate but complementary. The grammar performs that part of ellipsis resolution which is linguistically governed — it produces a representation which is underspecified by abstraction but which is constrained in the way it can be fully resolved by use of the relevant linguistic features — but needs no knowledge of context or inference. The DME (specifically the grounding process) then provides possible contextual values which can be used in resolution.

This approach also seems to offer a tidier solution than that of (Schlangen, 2003): in his approach, the semantic content of an elliptical fragment is defined around a fully specified but unknown relation *unknown_rel*. This is taken to denote the set of all possible relations that can concern that fragment (for the fragment *"John"*, the infinite set of sentences { *"John snores"*, *"John sleeps"*, *"John likes Mary"*, . . . , *"Bill has never been sky-diving with John before 7:25 on a Tuesday morning"*, . . . }. Instead, under the abstraction approach a fragment denotes a single semantic object, an abstract which can be thought of as $\lambda P.P(john)$, where P corresponds to the propositional content of the MAX-QUD question (the abstract is actually more complex than this – SAL-UTT and the other C-PARAMS are also abstracted). It also avoids postulating a new anaphoric *unknown_rel* relation as Schlangen's approach does: rather than have to replace or enrich this with a known relation in resolving the ellipsis, the abstracted parameters simply have to be instantiated in the process of applying the abstract to the context.

5.3.2   Contextual Clarification

It also takes care of the issue mentioned in section 5.2.3 above, that one might expect utterance-anaphoric elements to have to be grounded by identifying the previous utterance referent in context. As utterance-anaphoric phrases are defined as having their content identified with the value of SAL-UTT (see AVM (273) above), this now becomes the case, as the value of SAL-UTT is abstracted to C-PARAMS.

However, this raises the issue of whether it is possible to *clarify* the values of MAX-QUD and SAL-UTT. In the case of an utterance-anaphoric constituent, its content is associated with a contextual parameter (which must be identified with SAL-UTT), so it fits the definition of the coercion operations for clausal and constituent clarification, and we might therefore expect that CRs querying this parameter are possible. Similarly, a standard elliptical bare fragment's content is associated with a MAX-QUD contextual parameter, and we might expect this to be clarifiable.

Such CRs were not specifically identified in the empirical studies of chapter 3, but they do seem plausible. Example (280) shows an imagined example[6] querying SAL-UTT, where B's utterance intends to refer to a previous utterance of A's, but B's mishearing results in A being unable to identify the antecedent utterance:

(280)
A:	At least the pool was clean.
B:	What do you mean 'Mr Pool'?
A:	**Mr Pool?**
⤳	*"Which utterance of mine are you referring to by 'Mr Pool'?"*
⤳	*"Are you saying I said 'Mr Pool'?"*

Similarly, example (281) shows an equivalent for MAX-QUD. Here A cannot see the relevance of B's utterance (cannot find a question which it can be answering), and therefore cannot resolve the ellipsis. Note that the paraphrase given, *"What do you mean by 'Mary'?"*, asks about the meaning of the whole utterance, and is not the same as the question *"Who do you mean by Mary?"*, an equally plausible question, but one which asks about the identity of the individual referent named Mary, rather than the identity of the question under discussion. The two questions ask about different members of C-PARAMS: the former asks about the MAX-QUD parameter projected by the utterance as a whole, the latter the individual-referring parameter projected by the word *Mary*.

(281)
A:	I'm coming with you.
B:	Mary.
A:	**Mary?**
⤳	*"What do you mean by 'Mary'?"*
⤳	*"Which question is 'Mary' relevant to?"*

So it does seem reasonable to propose both MAX-QUD and SAL-UTT parameters, and that they behave as other members of C-PARAMS, requiring identification during grounding unless they are to lead to clarification. In fact, this approach goes some way towards explaining some of the CRs described in section 4.1.2 as *pragmatic* readings, which seem to be able to

---

[6]Example (280) is derived from a real BNC example, a reprise fragment CR with a lexical reading taken from file KPP, sentences 321–325:

(279)
Matthew:	It wasn't all that bad. At least the pool was clean.
Lara:	**Mr Pool?**
Matthew:	The pool.
Lara:	Oh <laugh>.

query the relevance of an utterance to the discourse. It's not clear that these can *only* query relevance, though, or indeed that all aspects of relevance can be explained via MAX-QUD and/or SAL-UTT.

5.3.3 Fragments in the Grammar

Taking this approach, the treatment of fragments in the grammar can otherwise follow that of SHARDS: the relations between semantic content, syntactic category and context being expressed via constraints on CONTEXT. The new abstraction of the contextual features then happens at *root-clause* level, as was the case with the abstraction of contextual information about speaker and addressee previously. This is expressed as the final version of the constraint on the *root-clause* type, as shown in AVM (282). As already sketched out in AVM (270), this specifies firstly how its illocutionary content is derived from the message associated with its head daughter; secondly that it does not merely inherit the C-PARAMS value of its daughter following the default constraint,[7] but adds members corresponding to speaker and addressee; and thirdly that we now also add members corresponding to MAX-QUD and SAL-UTT.

$$
(282) \quad
\begin{bmatrix}
\textit{root-cl} \\[4pt]
\text{CONT} \begin{bmatrix}
\textit{proposition} \\
\text{QUANTS} \quad \{\} \\[4pt]
\text{NUCL} \begin{bmatrix}
\textit{illoc-rel} \\
\text{UTT} & \boxed{i} \\
\text{ADD} & \boxed{j} \\
\text{MSG-ARG} & \boxed{1}
\end{bmatrix}
\end{bmatrix} \\[4pt]
\text{HEAD-DTR} \begin{bmatrix}
\text{CONT} & \boxed{1} \\
\text{C-PARAMS} & \boxed{2}
\end{bmatrix} \\[4pt]
\text{CTXT} \begin{bmatrix}
\text{SPKR} & \boxed{i} \\
\text{ADDR} & \boxed{j} \\
\text{MAX-QUD} & \boxed{q} \\
\text{SAL-UTT} & \boxed{s}
\end{bmatrix} \\[4pt]
\text{C-PARAMS} \left\{ \boxed{2} \cup \left\{
\begin{bmatrix} \boxed{i} : spkr(\boxed{i}) \end{bmatrix}, \begin{bmatrix} \boxed{j} : addr(\boxed{j}) \end{bmatrix}, \\
\begin{bmatrix} \boxed{q} : max\_qud(\boxed{q}) \end{bmatrix}, \begin{bmatrix} \boxed{s} : sal\_utt(\boxed{s}) \end{bmatrix}
\right\} \right\}
\end{bmatrix}
$$

The definitions of the fragment phrase and clause types can now follow SHARDS in general (the version given in (Fernández, 2002) with some of the extensions of (Dallas, 2001)), although some other changes will also be required: in relaxing the syntactic restriction to nominal categories (to allow clarification of signs with other categories including verbs, which we saw in chapter 3 to be possible although rare), and in the handling of STORE/QUANTS to

---

[7]As signs of type *root-clause* have only one daughter, the issue of inheritance vs. amalgamation discussed in chapter 4 is not an issue here.

reflect the changes to quantification described in chapter 4. The remaining part of this section takes the basic SHARDS definitions and shows how these modifications can be made.

Fragments in General

The grammatical treatment of fragments centres around the phrasal type *h(ea)d(ed)-frag(ment)-ph(rase)*, which specifies that fragments must be co-referential with, and exhibit syntactic parallelism with,[8] the contextually provided SAL-UTT (a salient constituent of a previous utterance).[9]

$$
(283) \quad
\begin{bmatrix}
\textit{hd-frag-ph} \\[4pt]
\text{HEAD} \quad
\begin{bmatrix}
\textit{verbal} \\
\text{VFORM} \quad \textit{fin}
\end{bmatrix} \\[10pt]
\text{HEAD-DTR} \quad
\begin{bmatrix}
\text{CAT} & \boxed{1} \\
\text{CONT} \,|\, \text{INDEX} & \boxed{2}
\end{bmatrix} \\[10pt]
\text{CTXT} \,|\, \text{SAL-UTT} \quad
\begin{bmatrix}
\text{CAT} & \boxed{1} \\
\text{CONT} \,|\, \text{INDEX} & \boxed{2}
\end{bmatrix}
\end{bmatrix}
$$

The restriction in SHARDS and G&C that this type only be applicable to NPs is removed – any head daughter whose content is a *parameter* will fit this constraint, and this now includes all content and function words as well as all NPs – in other words, all the categories that we have seen to allow reprise fragments. As discussed in chapter 4, reprises of other phrases such as VPs will be analysed as focussed reprises of their directly reprisable sub-constituents.

Short Answers

Declarative short answers have their properties defined by a subtype of *hd-frg-ph*, the phrase type *decl(arative)-frag(ment)-cl(ause)*, which specifies that their content is made up from the propositional content of the contextually provided MAX-QUD question, quantified over by the existentially quantified elements of both question and fragment. The version used here is modified to take into account the representation of QUANTS as a set of parameters, rather than

---

[8]As both Fernández et al. (2004a) and Schlangen (2003) point out, the requirement for strict syntactic parallelism may be too strong in some cases or for some categories. This is an issue for all fragments, not just CRs – for now, simple strict parallelism is assumed.

[9]Ginzburg et al. (2001a); Fernández (2002) take the type *hd-frg-ph* to apply only to fragments that constitute *arguments* (essentially NPs) rather than *adjuncts* (PPs and other modifiers), with the main differences being that bare adjuncts have a semantic content which is a modifier rather than a parameter, and can be used when there is no explicit antecedent. In the elliptical CRs we are concerned with here, the queried content is always a parameter (see chapter 4) and always has an antecedent source, so an argument-style analysis is all that is needed.

an ordered list of quantifiers (see chapter 4):

$$
(284) \quad
\begin{bmatrix}
\textit{decl-frag-cl} \\[4pt]
\text{CONT} & \begin{bmatrix}
\textit{proposition} \\
\text{SIT} & \boxed{1} \\
\text{SOA} & \begin{bmatrix} \text{QUANTS} & \boxed{Q_2} \uplus \boxed{Q_1} \\ \text{NUCL} & \boxed{2} \end{bmatrix}
\end{bmatrix} \\[4pt]
\text{STORE} & \boxed{Q_3} \\[4pt]
\text{HEAD-DTR} \,|\, \text{STORE} & \boxed{Q_2} \uplus \boxed{Q_3} \\[4pt]
\text{CTXT} & \begin{bmatrix}
\text{MAX-QUD} & \begin{bmatrix}
\textit{question} \\
\text{PARAMS} & \textit{non-empty-set} \\
\text{PROP} & \begin{bmatrix}
\textit{proposition} \\
\text{SIT} & \boxed{1} \\
\text{SOA} & \begin{bmatrix} \text{QUANTS} & \boxed{Q_1} \\ \text{NUCL} & \boxed{2} \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$
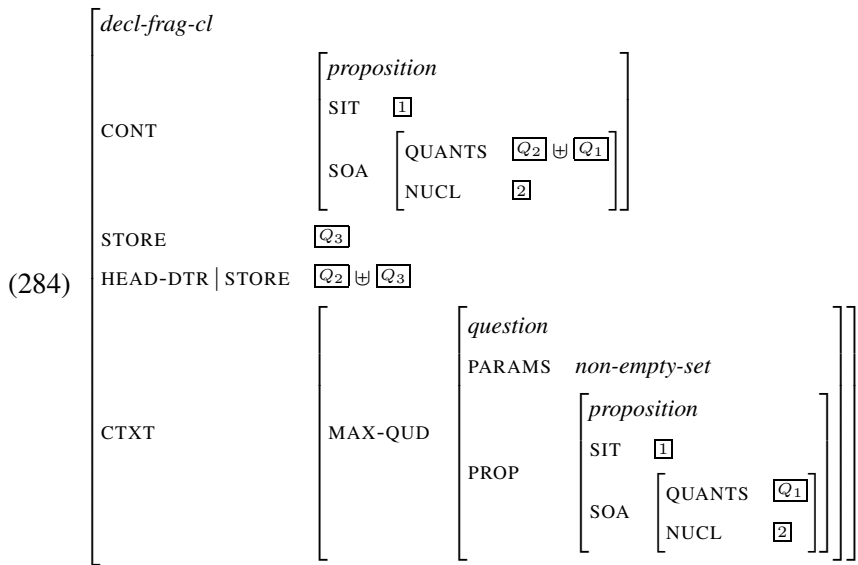
When combined with the constraints on *hd-frag-ph* (shown shaded ▓) and the general *root-cl* constraints (shown shaded ░), a standard declarative short answer *"John"* would therefore be given a representation as in AVM (285):[10]

$$
(285) \quad
\begin{bmatrix}
\text{PHON} & \langle \text{john} \rangle \\[4pt]
\text{CONT} & \begin{bmatrix}
\text{SIT} & \boxed{1} \\
\text{SOA} & \begin{bmatrix} \text{QUANTS} & \boxed{Q_1} \\ \text{NUCL} & \boxed{2} \end{bmatrix}
\end{bmatrix} \\[4pt]
\text{STORE} & \{\} \\[4pt]
\text{HEAD-DTR} & \begin{bmatrix}
\text{STORE} & \{\} \\
\text{CAT} & \boxed{3} \\
\text{CONT} & \boxed{7}\big[\,\boxed{4} : name(\boxed{4}, john)\big] \\
\text{C-PARAMS} & \{ \boxed{7} \}
\end{bmatrix} \\[4pt]
\text{CTXT} & \begin{bmatrix}
\text{MAX-QUD} & \boxed{5}\begin{bmatrix}
\text{PARAMS} & \textit{non-empty-set} \\
\text{PROP} & \begin{bmatrix}
\text{SIT} & \boxed{1} \\
\text{SOA} & \begin{bmatrix} \text{QUANTS} & \boxed{Q_1} \\ \text{NUCL} & \boxed{2} \end{bmatrix}
\end{bmatrix}
\end{bmatrix} \\
\text{SAL-UTT} & \boxed{6}\begin{bmatrix} \text{CAT} & \boxed{3} \\ \text{CONT}\,|\,\text{INDEX} & \boxed{4} \end{bmatrix}
\end{bmatrix} \\[4pt]
\text{C-PARAMS} & \{ \boxed{5}, \boxed{6}, \boxed{7} \}
\end{bmatrix}
$$

Less formally (without showing situations, quantification or syntactic parallelism con-

---

[10]AVM (285) omits the level of illocutionary force in CONTENT associated with root clauses, for ease of reading. It should be assumed in this and subsequent AVMs in this section.

straints) we can abbreviate this as follows:

$$(286) \quad \begin{bmatrix} \text{PHON} & \langle \text{john} \rangle \\ \text{CONT} \mid \text{SOA} \mid \text{NUCL} & \boxed{2} \\ \text{HEAD-DTR} \mid \text{CONT} & \boxed{7}\left[\, \boxed{4} : name(\boxed{4}, john)\right] \\ \text{CTXT} & \begin{bmatrix} \text{MAX-QUD} & \boxed{5}\left[?\{\dots\}.\boxed{2}\right] \\ \text{SAL-UTT} & \boxed{6}\left[\text{CONT} \quad \left[\boxed{4}:\dots\right]\right] \end{bmatrix} \\ \text{C-PARAMS} & \left\{\boxed{5}, \boxed{6}, \boxed{7}\right\} \end{bmatrix}$$

During grounding, this abstract is applied to the context: given a contextually available MAX-QUD *"Who likes Mary?"*, its corresponding SAL-UTT *"Who"*, and a contextually available referent for *"John"*, the members of C-PARAMS are instantiated to these values and give the fully specified declarative content:

$$(287) \quad \begin{bmatrix} \text{PHON} & \langle \text{john} \rangle \\ \text{CONT} \mid \text{SOA} \mid \text{NUCL} & \boxed{2} \\ \text{HEAD-DTR} \mid \text{CONT} & \left[\boxed{4}:name(\boxed{4}, john)\right] \\ \text{CTXT} & \begin{bmatrix} \text{MAX-QUD} & \left[?\boxed{5}.\boxed{2}like(\boxed{4}, mary)\right] \\ \text{SAL-UTT} & \begin{bmatrix} \text{PHON} & \langle \text{who} \rangle \\ \text{CONT} & \boxed{5}\left[\boxed{4}:person(\boxed{4})\right] \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Short Interrogatives

Direct (non-reprise) interrogative equivalents (*"John?"* ⤳ *"Is it John that likes Mary?"*) are produced in entirely parallel fashion, using an interrogative clause type (*dir(ect)-i(n)s(itu)-int(errogative)-cl(ause)* – see G&C) which embeds the propositional content of a declarative daughter within a question, as shown in AVM (288):

$$(288) \quad \begin{bmatrix} \textit{dir-is-int-cl} \\ \text{CONT} & \begin{bmatrix} \textit{question} \\ \text{PARAMS} & \Sigma_2 \\ \text{PROP} & \boxed{1} \end{bmatrix} \\ \text{STORE} & \Sigma_1 \\ \text{HEAD-DTR} & \begin{bmatrix} \text{CONT} & \boxed{1} \\ \text{STORE} & \left\{\Sigma_1 \uplus \Sigma_2\right\} \end{bmatrix} \end{bmatrix}$$

By taking *dir-is-int-cl* as the head daughter of the top-level *root-cl*, and the mother of a *decl-frag-cl*, this declarative fragment's proposition is used to form a question; for *"John?"* this results in a polar question, and a representation as in AVM (289) (again, constraints

inherited from the *hd-frag-ph* and *root-cl* types are shown shaded).

(289)

$$
\begin{bmatrix}
\text{PHON} & \langle \text{john} \rangle \\[4pt]
\text{CONT} & \begin{bmatrix} ?.\boxed{1} \end{bmatrix} \\[8pt]
\text{HEAD-DTR} & \begin{bmatrix}
\textit{decl-frag-cl} \\[4pt]
\text{CONT} & \boxed{1}\begin{bmatrix} \text{SOA} \mid \text{NUCL} & \boxed{2} \end{bmatrix} \\[4pt]
\text{HEAD-DTR} \mid \text{CONT} & \boxed{7}\begin{bmatrix} \boxed{4} : name(\,\boxed{4}\,, john) \end{bmatrix}
\end{bmatrix} \\[20pt]
\text{CTXT} & \begin{bmatrix}
\text{MAX-QUD} & \boxed{5}\begin{bmatrix} ?\{\ldots\}.\boxed{2} \end{bmatrix} \\[4pt]
\text{SAL-UTT} & \boxed{6}\begin{bmatrix} \text{CONT} & \begin{bmatrix} \boxed{4}:\ldots \end{bmatrix} \end{bmatrix}
\end{bmatrix} \\[16pt]
\text{C-PARAMS} & \left\{ \boxed{5}, \boxed{6}, \boxed{7} \right\}
\end{bmatrix}
$$
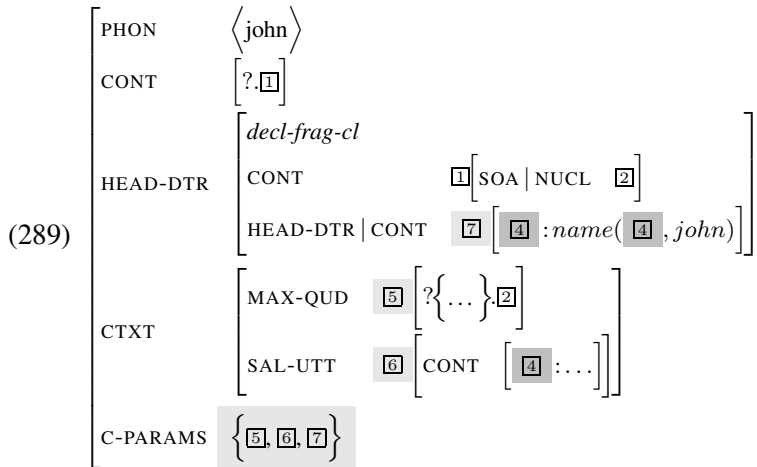
Direct Sluices

The third type of fragment treated by SHARDS is the direct (non-reprise) sluice. For this, a different subtype of *hd-frg-ph* is used, *sluice-interrogative-clause*, which specifies that the (interrogative) content is made up from the content of the contextually provided MAX-QUD question with the addition of a new *wh*-parameter contributed by the bare *wh*-phrase, while removing the widest-scoping non-negative quantifier (which is assumed to be associated with the *wh*-parameter index value by the pragmatic operations that govern MAX-QUD and SAL-UTT determination in SHARDS).

Given the new representation of quantification, there is no direct representation of scope in the QUANTS feature and thus no way of directly identifying a widest-scoping quantifier. Instead, the relevant quantified parameter can be directly identified. The index of the head daughter (the bare *wh*-phrase) will already be constrained to be co-referential with SAL-UTT, by virtue of the *hd-frg-ph* type. The content of SAL-UTT can therefore simply be constrained to be unified with *any* existentially quantified parameter in MAX-QUD, and this is removed from the QUANTS list of the overall clause.

The non-negative constraint on the quantifier in the original version is intended to prevent sluicing of sentences such as *"No man walks"*. In that version, this is achieved by typing: the *quant-rel* type (the semantic type for quantifiers) is divided into negative and positive subtypes (*no* being negative), and the quantifier associated with the sluice is constrained to be positive. Given the proposed analysis here of monotone-decreasing quantifiers via pairs of reference and complement set, a phrase like *"no man"* will now be given an analysis as shown in AVM (290), whereby the reference set r is empty and the complement set c is the

set of all men, with both of these sets being existentially quantified:

(290)
$$
\begin{bmatrix}
\text{PHON} & \left\langle \text{no, man} \right\rangle \\[2ex]
\text{CONTENT} & \begin{bmatrix} \text{REF} & \boxed{1}\left[r:r=Q(P)=\emptyset\right] \\[1ex] \text{COMP} & \boxed{2}\left[c:c=(P-r)=P\right] \end{bmatrix} \\[3ex]
\text{STORE} & \left\{\boxed{1},\boxed{2}\right\} \\[2ex]
\text{C-PARAMS} & \left\{\left[Q:Q=no''\right],\left[P:name(P,man)\right]\right\}
\end{bmatrix}
$$

Sluices must now be prevented from taking as antecedent either the empty reference set or the non-empty complement set. Indeed, it seems that sluices cannot ask about complement sets in general – in example (291), the sluice can only refer to the reference set of those who *did* go to the party, rather than those who *didn't*:

(291)

    A: Few of the students went to the party.
    B: Who?

    $\rightsquigarrow$ *"Which students* did *go to the party?"*
    $\rightsquigarrow$ #*"Which students* didn't *go to the party?"*

So the desired restriction can be achieved here by assuming that the *index* type (the semantic type for INDEX values) is classified according to two independent dimensions: in one, it is split into the subtypes *empty* and *non-empty*, where the empty set is the only object with *empty* type; in the other, into the subtypes *refset* and *compset*. The definition for reference-complement set pairs would of course ensure that the types in the latter dimension are assigned according to a parameter's role in the pair:

(292)
$$
\begin{bmatrix}
\textit{ref-comp} \\[1ex]
\text{REF} & \begin{bmatrix} \textit{parameter} \\ \text{INDEX} & \textit{refset} \end{bmatrix} \\[3ex]
\text{COMP} & \begin{bmatrix} \textit{parameter} \\ \text{INDEX} & \textit{compset} \end{bmatrix}
\end{bmatrix}
$$

This latter distinction may seem *ad hoc*, but it may also be motivated by independent reasons: as Nouwen (2003) points out, reference set and complement set anaphora behave differently: there is a general preference for reference set reference, and the licensing of complement set reference seems to be subject to a constraint of the ability to infer non-emptiness (with this constraint not applying to reference set anaphora). It seems likely that this type distinction (or at least a distinction along similar lines) will be required for a full account of anaphora.

Given this, the INDEX value of the quantified parameter associated with the sluice can

now be specified as *non-empty* and *refset*, as shown in AVM (293).

(293)

$$
\begin{bmatrix}
\textit{slu-int-cl} \\[4pt]
\text{CONT}
\begin{bmatrix}
\textit{question} \\
\text{PARAMS} \quad \boxed{P} \\[4pt]
\text{PROP}
\begin{bmatrix}
\textit{proposition} \\
\text{SIT} \quad \boxed{1} \\[4pt]
\text{SOA}
\begin{bmatrix}
\text{QUANTS} \quad \boxed{Q} \\
\text{NUCL} \quad \boxed{2}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix} \\[4pt]
\text{STORE} \quad \{\} \\[4pt]
\text{HEAD-DTR} \quad \begin{bmatrix} \text{STORE} \quad \boxed{P}\ \textit{set(wh-param)} \end{bmatrix} \\[4pt]
\text{CTXT}
\begin{bmatrix}
\text{MAX-QUD}
\begin{bmatrix}
\textit{question} \\
\text{PARAMS} \quad \{\} \\[4pt]
\text{PROP}
\begin{bmatrix}
\textit{proposition} \\
\text{SIT} \quad \boxed{1} \\[4pt]
\text{SOA}
\begin{bmatrix}
\text{QUANTS} \quad \left\{ \boxed{3}\begin{bmatrix} \textit{parameter} \\ \text{INDEX} \quad \textit{non-empty \& refset} \end{bmatrix} \right\} \uplus \boxed{Q} \\[4pt]
\text{NUCL} \quad \boxed{2}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix} \\[4pt]
\text{SAL-UTT} \quad \begin{bmatrix} \text{CONT} \quad \boxed{3} \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

A bare *wh*-phrase *"Who?"* would therefore obtain an analysis as in AVM (294), with its resulting content being a question derived from the MAX-QUD question, with the queried parameter being co-referential with both SAL-UTT and the corresponding existentially quantified parameter. Note that as *wh*-phrases do not contribute their content to C-PARAMS, only the MAX-QUD and SAL-UTT parameters now remain:[11]

(294)

$$
\begin{bmatrix}
\text{PHON} \quad \langle \text{who} \rangle \\[4pt]
\text{CONT} \quad \begin{bmatrix} ?\boxed{7}.\boxed{2} \end{bmatrix} \\[4pt]
\text{HEAD-DTR} \mid \text{CONT} \quad \boxed{7}\begin{bmatrix} \boxed{3} : person(\boxed{3}) \end{bmatrix} \\[4pt]
\text{CTXT}
\begin{bmatrix}
\text{MAX-QUD} \quad \boxed{5}\begin{bmatrix} ?.\exists\{\boxed{4}\}.\boxed{2} \end{bmatrix} \\[4pt]
\text{SAL-UTT} \quad \boxed{6}\begin{bmatrix} \text{CONT} \quad \boxed{4}\begin{bmatrix} \boxed{3} : \ldots \end{bmatrix} \end{bmatrix}
\end{bmatrix} \\[4pt]
\text{C-PARAMS} \quad \{\boxed{5}, \boxed{6}\}
\end{bmatrix}
$$
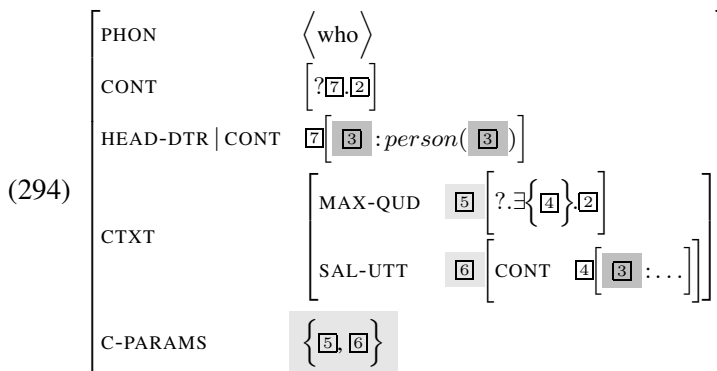
Of course, there is now no grammatical way of ensuring that only the widest-scoping quantifier in the antecedent utterance can be associated with the sluice. Actually, this seems quite natural – Lappin (2002) points out that there are examples in which sluices can be

---

[11]Actually, a contextual parameter corresponding to a predicate will be projected (for *"Who"*, the predicate *person* – see section 4.6.2), but not the entity-referring parameter which is the overall content of the *wh*-word.

associated with narrow-scoping quantifiers:

(295)

	A:	Each student will consult a supervisor.
	B:	Which one?
	$\rightsquigarrow$	*"Which supervisor will each student consult?"*

In the current analysis, the narrow-scoping parameter *"a supervisor"* would be functional on the wider-scoping parameter *"each student"*, but still available for sluicing, given a suitable functional analysis of the *wh*-phrase *"which one"* which could fit the constraints of AVM (293) above.[12]

5.4   A Grammar of CRs

This section now shows how the grammar can use the general features and the approach to fragments described above to implement and extend G&C's analysis to all the forms and readings that we have seen so far.

Contextual Coercion   All CR types will be analysed here as requiring some form of coercion operation to be fully understood. By their nature, CRs are asking about a previous source utterance or sub-utterance, which is not necessarily salient in the discourse before the CR appears. In order to understand the reference of CRs properly, a hearer must therefore be able to make this source utterance available; and for the elliptical reprise forms, she must also be able to make a salient question available to fully understand the content. This availability must be due to a pragmatic process: for Schlangen (2003) reasoning in a non-monotonic logic, for G&C a coercion operation which produces suitable SAL-UTT and MAX-QUD values. Here we use G&C's approach, but implemented via IS update rules as part of the dialogue system's DME; this must be assumed here, but will be explained fully in chapter 6 (sections 6.3 and 6.5). All CRs, then, will have an associated SAL-UTT contextual parameter which must be grounded to the source (sub-)utterance; reprise versions will also rely on a MAX-QUD parameter being grounded to a suitable CR question.

5.4.1   Non-Reprise CRs

As already stated in chapter 3, the analysis of non-reprise CRs can follow G&C now that we have an utterance-anaphoric grammar. Questions such as *"What did you say?"* are given a standard grammatical analysis, as sketched out in (87) and shown in detail here as (296). The verb *say* (along with other CR verbs such as *mean*) can be seen as taking the intended source utterance as an argument, thereby making the question one about that utterance's identity. This means *"What did you say?"* is in fact taken to mean *"What did you say [in that*

---

[12]A quantifier-based approach could also use a functional analysis to account for such examples – see G&S.

utterance]?"), which seems intuitively reasonable.

$$(296) \quad \begin{bmatrix} \text{PHON} & \langle \text{what, did, you, say} \rangle \\ \text{SLASH} & \{\} \\ \text{STORE} & \{\} \\ \text{CONT} & \begin{bmatrix} \textit{ask-rel} \\ \text{UTT} & \boxed{i} \\ \text{ADD} & \boxed{j} \\ \text{MSG-ARG} & \begin{bmatrix} \textit{question} \\ \text{PARAMS} & \{\boxed{1}\} \\ \text{PROP} & \boxed{2} \end{bmatrix} \end{bmatrix} \\ \text{CTXT} \mid \text{SAL-UTT} & \boxed{s} \\ \text{C-PARAMS} & \left\{ \begin{array}{l} \left[\boxed{s}: sal\_utt(\boxed{s})\right], \\ \left[\boxed{i}: spkr(\boxed{i})\right], \left[\boxed{j}: addr(\boxed{j})\right] \end{array} \right\} \end{bmatrix}$$

$$\begin{bmatrix} \text{PHON} & \langle \text{what} \rangle \\ \text{LOC} & \boxed{4}\left[\text{STORE} \quad \{\boxed{1}\}\right] \\ \text{CONT} & \boxed{1}\left[\text{INDEX} \quad \boxed{s}\right] \end{bmatrix}$$

$$\begin{bmatrix} \text{PHON} & \langle \text{did, you, say} \rangle \\ \text{SLASH} & \{\boxed{4}\} \\ \text{STORE} & \{\boxed{1}\} \\ \text{CONT} & \boxed{2}\begin{bmatrix} \textit{proposition} \\ \text{SOA} & \boxed{3} \end{bmatrix} \end{bmatrix}$$

$$\begin{bmatrix} \text{PHON} & \langle \text{did} \rangle \\ \text{SLASH} & \{\boxed{4}\} \\ \text{CONT} & \boxed{3} \end{bmatrix} \qquad \begin{bmatrix} \text{PHON} & \langle \text{you} \rangle \\ \text{STORE} & \{\} \\ \text{CONT} & \boxed{i}: spkr(\boxed{i}) \end{bmatrix} \qquad \begin{bmatrix} \text{PHON} & \langle \text{say} \rangle \\ \text{SLASH} & \{\boxed{4}\} \\ \text{CONT} & \boxed{3}\begin{bmatrix} \textit{utter-rel} \\ \text{SPKR} & \boxed{i} \\ \text{SIGN} & \boxed{s} \end{bmatrix} \\ \text{CTXT} \mid \text{SAL-UTT} & \boxed{s} \end{bmatrix}$$

    This means that a parameter for the queried source SAL-UTT becomes a member of C-PARAMS, and must be identified during grounding (establishing which utterance is being asked about). This will require a coercion operation to make the relevant utterance salient.

    Questions which are explicitly utterance-anaphoric (*"Who do you mean 'John'?"*, *"Did you say 'Mary'?"*) can now also be derived straightforwardly (as already shown in section 3.2.5): the phrases 'John' and 'Mary' can be given an utterance-anaphoric analysis by the grammar so that they can be taken as arguments by CR verbs which are defined to require

them (e.g. *say*, *mean*).

5.4.2 Conventional CRs

Just as the grammar specifies lexical entries for conventional phrases like greetings (section 5.2.2), it can specify content for conventional CRs. A conventional CR phrase *"Pardon?"* can be specified directly in the lexicon as having a content which corresponds to asking a suitable CR question, as shown in AVM (297) for lexical and AVM (298) for constituent versions. Clausal versions, as we saw, are not needed.

$$
(297)\quad
\begin{bmatrix}
\text{PHON} & \langle \text{pardon} \rangle \\[2pt]
\text{CONT} &
\begin{bmatrix}
\textit{ask-rel} \\
\text{UTT} & \boxed{i} \\
\text{ADD} & \boxed{j} \\
\text{MSG-ARG} &
\begin{bmatrix}
\textit{question} \\
\text{PARAMS} & \left\{ \boxed{s} \right\} \\
\text{PROP}\,|\,\text{SOA} &
\begin{bmatrix}
\textit{utter-rel} \\
\text{SPKR} & \boxed{j} \\
\text{SIGN} & \boxed{s}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix} \\[2pt]
\text{CTXT}\,|\,\text{SAL-UTT} & \boxed{s}\left[\textit{root-cl}\right] \\[2pt]
\text{C-PARAMS} & \left\{ \left[\boxed{s}:sal\_utt(\boxed{s})\right], \left[\boxed{i}:spkr(\boxed{i})\right], \left[\boxed{j}:addr(\boxed{j})\right] \right\}
\end{bmatrix}
$$

$$
(298)\quad
\begin{bmatrix}
\text{PHON} & \langle \text{pardon} \rangle \\[2pt]
\text{CONT} &
\begin{bmatrix}
\textit{ask-rel} \\
\text{UTT} & \boxed{i} \\
\text{ADD} & \boxed{j} \\
\text{MSG-ARG} &
\begin{bmatrix}
\textit{question} \\
\text{PARAMS} & \left\{ \boxed{1} \right\} \\
\text{PROP}\,|\,\text{SOA} &
\begin{bmatrix}
\textit{spkr-meaning-rel} \\
\text{SPKR} & \boxed{j} \\
\text{SIGN} & \boxed{s} \\
\text{CONT} & \boxed{1}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix} \\[2pt]
\text{CTXT}\,|\,\text{SAL-UTT} & \boxed{s}\left[\textit{root-cl}\right] \\[2pt]
\text{C-PARAMS} & \left\{ \left[\boxed{s}:sal\_utt(\boxed{s})\right], \left[\boxed{i}:spkr(\boxed{i})\right], \left[\boxed{j}:addr(\boxed{j})\right] \right\}
\end{bmatrix}
$$

In both cases, the constraint that the SAL-UTT must be of type *root-cl* ensures that the question expressed is asking about a complete utterance, rather than a sub-constituent (as was seen to be the case for conventional CRs in chapter 3).[13] The SAL-UTT must, as usual, be

---

[13]In fact, the same should probably hold for non-reprise *wh*-CRs such as *"What did you say?"*, *"What do you*

made a contextual parameter.

A more underspecified approach might be possible in which conventional CR phrases are specified as taking their contents entirely from a contextual MAX-QUD, as shown in AVM (299). This would then rely on coercion operations not only making the source utterance salient, but making a question about lexical identity or intended meaning salient – and these are exactly the operations we will rely on for reprises anyway. This would only be possible for conventional phrases which can take any CR reading (which might be possible – certainly phrases such as *"Eh?"* can take lexical or constituent readings). However, as there seems no principled way of ruling out unwanted readings (not only the clausal CR reading, but any other MAX-QUD question such as raised by asking a standard question) this is left aside for now as a future possibility.

$$
(299) \quad
\begin{bmatrix}
\text{PHON} & \langle \text{pardon} \rangle \\[4pt]
\text{CONT} &
\begin{bmatrix}
\textit{ask-rel} \\
\text{UTT} & \boxed{i} \\
\text{ADD} & \boxed{j} \\
\text{MSG-ARG} & \boxed{q}
\end{bmatrix} \\[10pt]
\text{CTXT} &
\begin{bmatrix}
\text{MAX-QUD} & \boxed{q} \\
\text{SAL-UTT} & \boxed{s}\,[\textit{root-cl}]
\end{bmatrix} \\[10pt]
\text{C-PARAMS} &
\left\{
\begin{array}{l}
\big[\boxed{q}\!:\!max\_qud(\boxed{q})\big],\ \big[\boxed{s}\!:\!sal\_utt(\boxed{s})\big], \\[4pt]
\big[\boxed{i}\!:\!spkr(\boxed{i})\big],\ \big[\boxed{j}\!:\!addr(\boxed{j})\big]
\end{array}
\right\}
\end{bmatrix}
$$

5.4.3 Clausal Reprises

Reprise CRs cannot have their content specified directly, and must rely on coercion operations to provide it via contextual parameters. Clausal reprises use the *parameter focussing* operation defined by G&C and modified to allow for sub-constituent focussing in section 4.5.4, shown there as (220) and repeated here as (300). The operation produces a question by abstracting a problematic parameter associated with the source utterance itself or one of its sub-constituents from its overall content:

$$
(300) \quad
\begin{bmatrix}
\text{C-PARAMS} & \{\ldots, \boxed{1}, \ldots\} \\[4pt]
\text{CONSTITS} & \left\{\ldots, \boxed{2}\Big[\text{CONSTITS} \ \{\ldots, [\text{CONTENT}\ \boxed{1}], \ldots\}\Big], \ldots\right\} \\[4pt]
\text{CONTENT} & \boxed{4}
\end{bmatrix} \quad \text{(original utterance)}
$$
$$
\Rightarrow
$$
$$
\begin{bmatrix}
\text{CONTEXT} &
\begin{bmatrix}
\text{SAL-UTT} & \boxed{2} \\
\text{MAX-QUD} & ?\boxed{1}.\boxed{4}
\end{bmatrix}
\end{bmatrix} \quad \text{(partial reprise context description)}
$$

mean?". This could be specified in a similar way, as a restriction on the utterance-referring version of the word *what*.

The analyses in this clausal section therefore all involve contextual parameters associated with MAX-QUD and SAL-UTT which must be grounded to values produced as above. Note that in the case of clausal CRs, there will also be contextual parameters associated with the standard semantic content of the CR (in a reprise fragment *"John?"*, the referent of *John*) – if any of these parameters cannot be identified, a CR-of-CR may ensue. For utterance-anaphoric versions (see below), these parameters from standard content may of course not be present.

Literal Reprises

As section 3.2 showed, this clausal reading is all that is required for literal (full-sentence) reprises. Following G&S and G&C, these sentences are analysed via a *reprise-int-cl* type which defines their content as coming from context, but we can now make the link with MAX-QUD coercion explicit (rather than using G&C's separate *prev-utt* relation), and similarly the link with focus:

(301)

$$
\begin{bmatrix}
\textit{reprise-int-cl} \\[4pt]
\text{CONT} \quad
\begin{bmatrix}
\textit{question} \\
\text{PARAMS} & \boxed{P_2} \\
\text{PROP} & \boxed{1}\left[\text{SOA}\,|\,\text{NUCL}\,|\,\text{MSG-ARG} \quad \boxed{2}\right]
\end{bmatrix} \\[12pt]
\text{STORE} \quad \boxed{P_1} \\[6pt]
\text{INFO-STRUCT} \quad
\begin{bmatrix}
\text{FOCUS} &
\begin{bmatrix}
\text{CAT} & \boxed{C_f} \\
\text{CONT}\,|\,\text{INDEX} & \boxed{I_f}
\end{bmatrix}
\end{bmatrix} \\[12pt]
\text{HEAD-DTR} \quad
\begin{bmatrix}
\text{CONT} & \boxed{2} \\
\text{STORE} & \boxed{P_1} \uplus \boxed{P_2}
\end{bmatrix} \\[12pt]
\text{CTXT} \quad
\begin{bmatrix}
\text{MAX-QUD} &
\begin{bmatrix}
\text{PARAMS} & \left\{\boxed{3}\right\} \\
\text{PROP} & \boxed{1}
\end{bmatrix} \\[12pt]
\text{SAL-UTT} &
\left[\text{CONSTITS} \quad \left\{\ldots\begin{bmatrix}\text{CAT} & \boxed{C_f} \\ \text{CONT} & \boxed{3}\left[\text{INDEX} \quad \boxed{I_f}\right]\end{bmatrix}\ldots\right\}\right]
\end{bmatrix}
\end{bmatrix}
$$
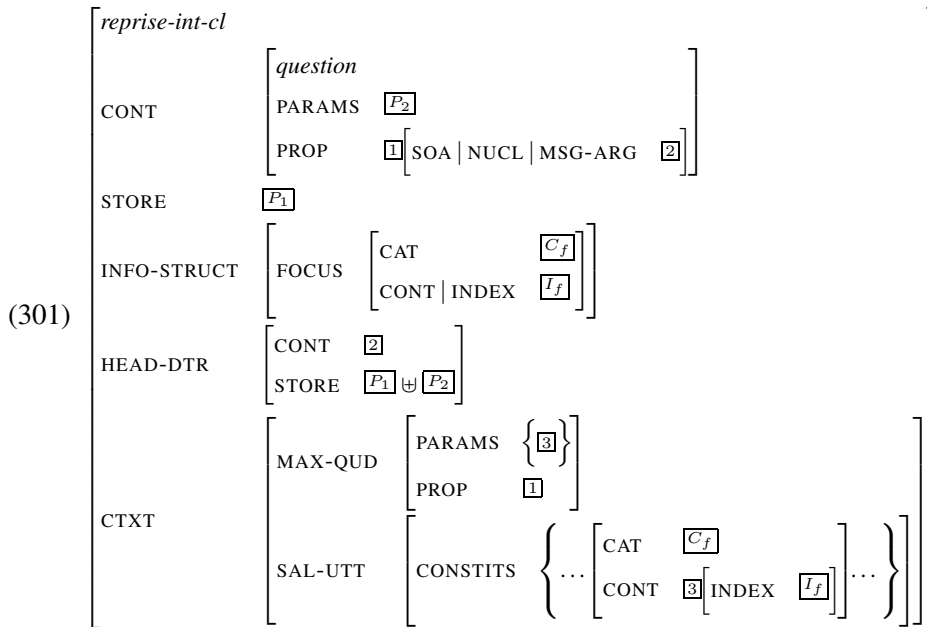
Firstly, this tells us that the content will be a clausal question (it asks about an illocutionary relation) which comes from MAX-QUD, and which is constrained to ask about the literal content of the reprise (its head daughter's content) – e.g. *"Does John like Mary?"* must be asking about a previous question about John liking Mary, rather than, say, Bill liking Sue. Secondly, it ensures that any focussed sub-constituent is parallel to the constituent of SAL-UTT which is queried by (abstracted in) MAX-QUD – e.g. in *"Does JOHN like Mary?"* requires the MAX-QUD to be the question *"Who$_i$ is it that you are asking whether i likes Mary?"*, rather than another clausal question produced by abstracting a different parameter. Once these contextual parameters have been produced by coercion and the reprise grounded

to them, the full interpretation will be as below:

(302)

$$
\begin{bmatrix}
\textit{reprise-int-cl} \\
\text{PHON} \quad \langle \text{does, JOHN, like, mary} \rangle \\[4pt]
\text{CONT} \quad \begin{bmatrix} ?.\boxed{1} \end{bmatrix} \\[6pt]
\text{INFO-STRUCT} \quad \begin{bmatrix} \text{FOCUS} \begin{bmatrix} \text{PHON} \quad \langle \text{JOHN} \rangle \\ \text{CONT | INDEX} \quad \boxed{I} \end{bmatrix} \end{bmatrix} \\[10pt]
\text{HEAD-DTR} \quad \begin{bmatrix} \text{CONT} \quad \boxed{2} \end{bmatrix} \\[8pt]
\text{CTXT} \quad \begin{bmatrix}
\text{MAX-QUD} \quad \begin{bmatrix} ?\{\boxed{3}\}.\boxed{1} \end{bmatrix} \\[8pt]
\text{SAL-UTT} \quad \begin{bmatrix}
\text{PHON} \quad \langle \text{does, john, like, mary} \rangle \\
\text{CONT} \quad \boxed{1}\begin{bmatrix} ask(i, j, \boxed{2}?.like(j, m)) \end{bmatrix} \\[8pt]
\text{CONSTITS} \quad \left\{ \ldots \begin{bmatrix} \text{PHON} \quad \langle \text{john} \rangle \\ \text{CONT} \quad \boxed{3}\begin{bmatrix} \boxed{I}\, j : name(\boxed{I}\,j, john) \end{bmatrix} \end{bmatrix} \ldots \right\}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

WH-Substituted Reprises

Clausal versions of *wh*-substituted reprise sentences follow directly from the definition above: the constraint on STORE in AVM (301) ensures that the *wh*-parameter introduced by the *wh*-phrase is made a member of PARAMS, thus making the overall question a *wh*-question rather than a polar question. As long as *wh*-phrases are seen as always being in focus (see section 4.5.4), the information structure constraint also still applies, making sure that the contextual MAX-QUD question is the correct one.

As long as a version of the *wh*-word *what* is defined in the lexicon which can have a predicate as its INDEX value (rather than an individual), it can be used in place of any phrase which refers to a predicate: this allows e.g. CNs to be *wh*-substituted too. However, for lexical readings, which seemed to be the most important, a small change is required – see section 5.4.4 below.

Clausal Reprise Fragments

The treatment of fragments (section 5.3.3, and specifically short interrogatives, gives all that is needed for clausal reprise fragment CRs: the analysis follows G&C and is exactly as in AVM (289) above – as long as coercion produces the appropriate clarification question QUD and salient utterance, resolution to give the CR content proceeds by grounding the relevant parameters. Once the contextual parameters have been resolved, a fully specified form might look as in AVM (303). Here the initial source utterance was something like *"Does John like Mary?"*, producing a possible MAX-QUD question *"Who$_i$ (named John) are you asking*

whether i likes Mary?" and a SAL-UTT utterance *"John"*. The clausal fragment *"John Smith?"* can then be resolved as below:

$$(303)\quad \begin{bmatrix} \text{PHON} & \left\langle \text{john, smith} \right\rangle \\ \text{CONT} & \begin{bmatrix} ?.\boxed{1} \end{bmatrix} \\ \text{STORE} & \{\} \\ \text{HEAD-DTR}\,|\,\text{CONT} & \begin{bmatrix} \boxed{2}:name(\boxed{2}, john\_smith) \end{bmatrix} \\ \text{CTXT} & \begin{bmatrix} \text{MAX-QUD} & \begin{bmatrix} ?\boxed{3}.\boxed{1}\begin{bmatrix} ask(i,j,like(\boxed{2}, mary)) \end{bmatrix} \end{bmatrix} \\ \text{SAL-UTT} & \begin{bmatrix} \text{CONT} & \boxed{3}\begin{bmatrix} \boxed{2}:name(\boxed{2}, john) \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Reprise Sluices

Reprise sluices (which only require a clausal analysis) are analysed along exactly the same lines, using the *decl-frag-cl* type with an interrogative mother to give an interrogative question, which is fully specified by resolution of the contextual features during grounding of C-PARAMS to values provided by a pragmatic operation. The only difference is that this time the content is a *wh*-question rather than a polar question, as the sluice fragment itself puts a *wh*-parameter into STORE, which is retrieved higher up as part of the question's PARAMS feature (again, see G&C for more detail).

$$(304)\quad \begin{bmatrix} \text{PHON} & \left\langle \text{who} \right\rangle \\ \text{CONT} & \begin{bmatrix} ?\boxed{4}.\boxed{1} \end{bmatrix} \\ \text{STORE} & \{\} \\ \text{HEAD-DTR} & \begin{bmatrix} \text{CONT} & \boxed{4}\begin{bmatrix} \boxed{2}:person(\boxed{2}) \end{bmatrix} \\ \text{STORE} & \left\{\boxed{4}\right\} \end{bmatrix} \\ \text{CTXT} & \begin{bmatrix} \text{MAX-QUD} & \begin{bmatrix} ?\boxed{3}.\boxed{1}\begin{bmatrix} ask(i,j,like(\boxed{2}, mary)) \end{bmatrix} \end{bmatrix} \\ \text{SAL-UTT} & \begin{bmatrix} \text{CONT} & \boxed{3}\begin{bmatrix} \boxed{2}:name(\boxed{2}, john) \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

There is nothing in this analysis preventing reprise sluices from being given other CR readings if the pragmatic coercion operation can produce a suitable MAX-QUD and SAL-UTT (and indeed G&C show how a constituent reprise sluice can be constructed along these lines). However, chapter 3 suggested that only clausal readings were required for sluices: in chapter 6 (section 6.3) we will see how the coercion operations can be restricted to prevent this ambiguity if desired.

5.4.4 Utterance-Anaphoric Reprises

As already outlined, both the constituent and lexical readings will require an utterance-anaphoric treatment. The concept and definition of the utterance-anaphoric phrase type *utt-anaph-ph* has already been introduced in section 5.2.3 above: here it is applied to elliptical fragments to give the remaining required readings of reprise fragments, together with an analysis of gaps and fillers.

Lexical Reprise Fragments

The standard short interrogative (*dir-is-int-cl* with *decl-frag-cl*) fragment analysis above is now almost enough to give a suitable analysis for lexical readings, when combined with utterance-anaphora. However, as the standard *decl-frag-cl* type is a subtype of *hd-frag-ph*, it forces the referent of its head daughter to be identified with the content of the salient utterance – instead, we now want it to be identified with the salient utterance itself. We therefore need a new phrase type *utt-frag-ph* which requires an *utt-anaph-ph* as its daughter (which will therefore already be identified with SAL-UTT):

$$(305) \quad \begin{bmatrix} utt\text{-}frag\text{-}ph \\ \text{HEAD} & \begin{bmatrix} verbal \\ \text{VFORM} & fin \end{bmatrix} \\ \text{HEAD-DTR} & \begin{bmatrix} utt\text{-}anaph\text{-}ph \end{bmatrix} \end{bmatrix}$$

We can now define *utt-decl-frag-cl*, which is a subtype of *utt-frag-ph* and has exactly the same additional constraints as as *decl-frag-cl* above (see AVM (284)). We also require the *lexical identification* coercion operation (introduced in chapter 3 and shown here as AVM (306)), which produces a context in which the maximal QUD is a question about what word was uttered, and the salient utterance is the problematic utterance *U* itself:

$$(306) \quad \begin{aligned} & \begin{bmatrix} \text{CONSTITS} & \left\{ \dots, \boxed{1}, \dots \right\} \end{bmatrix} & \text{(original utterance)} \\ \Rightarrow \\ & \begin{bmatrix} \text{CONTEXT} & \begin{bmatrix} \text{SAL-UTT} & \boxed{1} \\ \text{MAX-QUD} & ?\boxed{1}.utter\_rel(a, \boxed{1}) \end{bmatrix} \end{bmatrix} & \text{(partial reprise context description)} \end{aligned}$$

Note that this operation is not currently defined to allow sub-constituent focussing, as we have seen no evidence that it is required – it could be added along the lines of the parameter focussing operation in AVM (300). Given these two definitions, the correct reading *"did you*

utter U?" is now obtained for a lexical reprise fragment – see AVM (307).

$$(307) \quad \begin{bmatrix} \text{PHON} & \boxed{2}\langle \text{john} \rangle \\ \text{CONT} & \begin{bmatrix} ?.\boxed{1} \end{bmatrix} \\ \text{HEAD-DTR} & \begin{bmatrix} \textit{utt-anaph-ph} \\ \text{CONT} & \boxed{3} \end{bmatrix} \\ \text{CTXT} & \begin{bmatrix} \text{MAX-QUD} & \begin{bmatrix} ?\boxed{3}.\boxed{1} \begin{bmatrix} utter(i,\boxed{3}) \end{bmatrix} \end{bmatrix} \\ \text{SAL-UTT} & \boxed{3}\begin{bmatrix} \text{PHON} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Sluices (and *wh*-substituted sentences) can receive a lexical analysis in the same way if desired, using the same approach as for clausal versions (but now treating the *wh*-word as a daughter of *utt-decl-frag-cl*). All that is required is a definition of an utterance-anaphoric *wh*-word, which refers to (but is not constrained to be phonologically parallel with) the SAL-UTT:

$$(308) \quad \begin{bmatrix} \textit{wh-utt-anaph-ph} \\ \text{CONT} & \begin{bmatrix} \textit{wh-param} \\ \text{INDEX} & \boxed{2} \end{bmatrix} \\ \text{CTXT} \mid \text{SAL-UTT} & \boxed{2} \end{bmatrix}$$

Sub-Lexical Queries   Sub-lexical *wh*-questions (e.g. *"What-jacency?"* – see section 2.3.2) can also use this analysis, provided that *wh*-substituted words can be parsed suitably. A lexical entry would be required along the lines of AVM (309):

$$(309) \quad \begin{bmatrix} \text{PHON} & \langle \text{what}, \boxed{3} \rangle \\ \text{CONT} & \boxed{1}\begin{bmatrix} \textit{wh-param} \\ \text{CONT} & \boxed{2}\begin{bmatrix} \text{PHON} & \langle \ldots, \boxed{3} \rangle \end{bmatrix} \end{bmatrix} \\ \text{STORE} & \{\boxed{1}\} \end{bmatrix}$$

Allowing such a *wh*-word to be a daughter of *utt-decl-frag-cl* would then produce some-

thing like AVM (310), once resolved with a coerced lexical CR MAX-QUD:

$$
(310) \quad
\begin{bmatrix}
\text{PHON} & \left\langle \text{what,jacency} \right\rangle \\
\text{CONT} & \left[\text{?}\boxed{3}.\boxed{1} \right] \\
\text{STORE} & \{\} \\
\text{HEAD-DTR} & \begin{bmatrix} \text{CONT} & \boxed{3} \\ \text{STORE} & \left\{ \boxed{3} \right\} \end{bmatrix} \\
\text{CTXT} & \begin{bmatrix} \text{MAX-QUD} & \left[\text{?}\boxed{3}.\boxed{1}\left[utter(i,\boxed{3}) \right] \right] \\ \text{SAL-UTT} & \boxed{3}\left[\text{PHON} \quad \left\langle \ldots, \text{jacency} \right\rangle \right] \end{bmatrix}
\end{bmatrix}
$$

This treatment differs from that of (Artstein, 2002) – *wh*-substituted words are not seen as denoting functions from partial strings (possible replacements for the *wh*-section) to words, but simply as referring to words, with a partial constraint on the referent word's form. This seems both more consistent with the approach so far, and more parsimonious.

Constituent Reprise Fragments

An analysis for constituent fragments is already provided by G&C, and this can be used without modification. The phrasal type *constit-clar-int-cl* (which can now be considered as a subtype of the new *utt-frag-ph* is used, which identifies fragment content with MAX-QUD directly:

$$
(311) \quad
\begin{bmatrix}
\textit{constit-clar-int-cl} \\
\text{CONT} & \boxed{1} \\
\text{HEAD-DTR} & \left[\textit{utt-anaph-ph} \right] \\
\text{CTXT} \mid \text{MAX-QUD} & \boxed{1}
\end{bmatrix}
$$

Given the definition of the *parameter identification* coercion operation (the version modified for sub-constituent focussing in section 4.5.4 and shown there as AVM (224) is used, repeated here as AVM (312)), the correct content is derived directly.

$$
(312) \quad
\begin{bmatrix}
\text{C-PARAMS} & \left\{ \ldots, \boxed{1}, \ldots \right\} \\
\text{CONSTITS} & \left\{ \ldots, \boxed{2}\left[\text{CONSTITS} \quad \left\{ \ldots, \boxed{3}\left[\text{CONTENT} \quad \boxed{1} \right], \ldots \right\} \right], \ldots \right\}
\end{bmatrix}
\quad \text{(original utterance)}
$$

$$
\Rightarrow
$$

$$
\begin{bmatrix}
\text{CONTEXT} & \begin{bmatrix} \text{SAL-UTT} & \boxed{2} \\ \text{MAX-QUD} & \text{?}\boxed{1}.spkr\_meaning\_rel(a,\boxed{3},\boxed{1}) \end{bmatrix}
\end{bmatrix}
\quad \begin{array}{c}\text{(partial reprise context}\\\text{description)}\end{array}
$$

However, this is not the only coercion operation available, and there is nothing to prevent the *constit-clar-int-cl* type from combining with any MAX-QUD question, including a clausal

question produced by the parameter focussing operation. As shown in AVM (313) then, for this particular CR form and reading an extra constraint can be added which requires that the MAX-QUD question be a constituent CR question – this prevents undesired readings from arising. Rather than applying this constraint to *constit-clar-int-cl* directly, it is expressed via a new subtype specifically for constituent fragments (this allows a further subtype to be defined for gaps below). Constraints inherited from the supertype are shown shaded:

$$(313) \quad \begin{bmatrix} \textit{frg-constit-clar-int-cl} \\ \text{CONT} & \boxed{1} \\ \text{HEAD-DTR} & \begin{bmatrix} \textit{utt-anaph-ph} \end{bmatrix} \\ \text{CTXT} \mid \text{MAX-QUD} & \boxed{1} \begin{bmatrix} \text{PROP} \mid \text{SOA} & \begin{bmatrix} \textit{spkr\_meaning\_rel} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Given a context in which the maximal QUD is a question about what the intended content of an antecedent utterance was, the correct reading is now derived (314). Note that the *decl-frag-cl* analysis used above cannot be used here, as we are essentially forming a *wh*-question without any constituent putting a *wh*-parameter into STORE – hence the requirement for the specific *constit-clar-int-cl* type.

$$(314) \quad \begin{bmatrix} \text{PHON} & \boxed{2} \langle \text{john} \rangle \\ \text{CONT} & \boxed{1} \\ \text{HEAD-DTR} & \begin{bmatrix} \textit{utt-anaph-ph} \\ \text{CONT} & \boxed{3} \end{bmatrix} \\ \text{CTXT} & \begin{bmatrix} \text{MAX-QUD} & \boxed{1} ?\boxed{4} \begin{bmatrix} \textit{spkr\_meaning\_rel}(i, \boxed{3}, \boxed{4}) \end{bmatrix} \\ \text{SAL-UTT} & \boxed{3} \begin{bmatrix} \text{PHON} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Reprise Gaps

Gaps are similar to constituent fragments in that they essentially ask a *wh*-question without using a *wh*-word. An analysis of reprise gaps can therefore use exactly the same grammatical construction (*constit-clar-int-cl*), but a different subtype distinguished by constraining the contextual MAX-QUD question to be different: a question about what utterance was next to be uttered, *"What word Y did you say after you said X?"*, where *X* is the SAL-UTT (and therefore the direct content of the fragment):

$$(315) \quad \begin{bmatrix} \textit{gap-constit-clar-int-cl} \\ \text{CONT} & \boxed{1} \\ \text{HEAD-DTR} & \begin{bmatrix} \textit{utt-anaph-ph} \end{bmatrix} \\ \text{CTXT} \mid \text{MAX-QUD} & \boxed{1} \begin{bmatrix} \text{PROP} \mid \text{SOA} & \begin{bmatrix} \textit{utter\_consec} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

When combined with a suitable context, and with the constraints of *utt-anaph-ph* on the

daughter, the reprise gap will be resolved as follows:

$$(316) \quad \begin{bmatrix} \text{PHON} & \boxed{2}\left\langle \text{the} \right\rangle \\ \text{CONT} & \boxed{1} \\ \text{HEAD-DTR}\,|\,\text{CONT} & \boxed{3} \\ \\ \text{CTXT} & \begin{bmatrix} \text{MAX-QUD} & \boxed{1}\left[?\boxed{4}.\left[utter\_consec(i,\boxed{3},\boxed{4}) \right] \right] \\ \text{SAL-UTT} & \boxed{3}\left[\text{PHON} \quad \boxed{2} \right] \end{bmatrix} \end{bmatrix}$$

This question must of course be produced by a new contextual coercion operation, termed *gap identification*, which is defined as part of the grounding process and given below in AVM (317).[14] As for the lexical identification operation, this definition does not allow any sub-constituent focussing, which seems correct for gaps. While the version here gives a lexical reading (which was all that was seen in chapter 3), others (e.g. constituent) would be possible by defining different coercion operations along the same lines.

$$(317) \quad \begin{bmatrix} \text{C-PARAMS} & \left\{ \ldots, \boxed{1}, \ldots \right\} \\ \\ \text{CONSTITS} & \left\{ \ldots, \boxed{3}\left[\text{PHON} \quad \boxed{2} \right], \boxed{4}\begin{bmatrix} \text{PHON} & \boxed{5} \\ \text{CONTENT} & \boxed{1} \end{bmatrix}, \ldots \right\} \\ \\ \text{PHON} & \left\langle \ldots, \boxed{2}, \boxed{5}, \ldots \right\rangle \end{bmatrix} \quad \text{(original utterance)}$$

$$\Rightarrow$$

$$\begin{bmatrix} \text{CONTEXT} & \begin{bmatrix} \text{SAL-UTT} & \boxed{3} \\ \text{MAX-QUD} & ?\boxed{4}.\left[utter\_consec(i,\boxed{3},\boxed{4}) \right] \end{bmatrix} \end{bmatrix} \quad \text{(partial reprise context description)}$$

Gap Fillers

Fillers appear more like standard lexical reprise fragments: they offer a word and ask a polar question (*"Is it this word that you are intending to utter next?"*). They can therefore use a similar analysis, using the utterance-anaphoric *utt-decl-frag-cl*. There are two differences: firstly, the coercion mechanism assumed to produce MAX-QUD and SAL-UTT must be different (producing the question about next intended utterance); secondly, the SAL-UTT which is asked about (and denoted anaphorically) does not actually exist in context yet, and its (intended) presence must be deduced given that the previous utterance was unfinished. This

---

[14]The restriction that the two constituents be consecutive in the original utterance is expressed here via consecutive PHON membership. Other ways of expressing this might be possible, for example via membership of DTRS, but the use of PHON seems directly motivated as the question relates to the order in the surface form of the utterance.

seems reasonable, and such an analysis would produce a result as in AVM (318) below.

$$(318) \quad \begin{bmatrix} \text{PHON} & \boxed{2}\langle \text{john} \rangle \\ \text{CONT} & \begin{bmatrix} ?.\boxed{1} \end{bmatrix} \\ \text{HEAD-DTR} \mid \text{CONT} & \boxed{3} \\ \\ \text{CTXT} & \begin{bmatrix} \text{MAX-QUD} & \begin{bmatrix} ?\boxed{3}.\boxed{1} \begin{bmatrix} intend(i, utter\_consec(i, \boxed{3})) \end{bmatrix} \end{bmatrix} \\ \text{SAL-UTT} & \boxed{3} \begin{bmatrix} \text{PHON} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Again, only the lexical reading seems to be required, as given above, although this kind of analysis would not preclude others. However, this has not been implemented in the grammar described here (or in the system of chapter 6) – as unfinished utterances cannot currently be parsed, there is no use for an analysis of fillers.

5.4.5 Corrections

Although it would be possible, corrections have not been implemented in the way suggested in chapter 3, which gave them the standard clausal, constituent or lexical readings further embedded within a proposition concerning intent. Instead, standard clausal and lexical fragments can be used to serve as corrections.

Such fragments would lack the meaning of explicitly querying intent that seems intuitively to be part of the correction reading: given a source utterance *"Did Bo leave?"*, a clausal fragment *"Jo?"* would mean *"Did you ask whether JO left? (rather than BO)"*, as opposed to *"Did you **intend** to ask whether JO left? (rather than BO)"*. However, this distinction does not seem crucial from a dialogue system perspective: the subsequent response will not be affected, as it will presumably be negative if Bo was intended, and affirmative if Jo was intended, in both cases. This will hold for CRs generated by both user and system.

Perhaps a level of fine-grained distinction is lost, but the standard readings do query the form or intended meaning of the source utterance, and as long as they need not always correspond to the *correct* original form or intended meaning, they will still allow it to be corrected or contradicted. An example of such contradiction and belief revision, involving a standard clausal reprise fragment, is given in section 6.5. Of course, it is also worth remembering that corrections seem to be very rare compared to other CRs, so such a fine-grained distinction may not be a significant loss anyway.

Given this, it seems reasonable to propose that an analysis which treats corrections as standard clausal CRs will be sufficient – in other words, that no separate analysis for corrections is really needed. Of course, this also has the advantage of not requiring the grammar to explicitly assign both types of reading and not requiring a dialogue system to disambiguate between them, thus reducing the complexity of the system.

5.5   Ambiguity

The grammar is ambiguous in that most sentences will have more than one possible interpretation: at least the standard interpretation (where one exists) and an utterance-anaphoric version, and more in cases of structural ambiguity. The parser (a simple bottom-up left-to-right chart parser taken from the SHARDS system) produces the set of all signs that correspond to the longest inactive edge in the chart (i.e. that cover the longest continuous parsable substring),[15] subject to the constraint that this is not a singleton set containing only an utterance-anaphoric edge of length greater than 1. This constraint is required as utterance-anaphoric phrases can be built from any string, and therefore any sentence (whether otherwise grammatical or not) can be given an utterance-anaphoric parse covering its full length: this would prevent other legitimate parses of less than full length being considered if an unconstrained longest-edge approach was taken.

Decisions about resolving ambiguity are left to the grounding process, where all contextual sources of information are available to help with disambiguation, rather than attempting to use the grammar (this is discussed more fully in chapter 6 where the grounding process is defined). For example, the grammar allows both function words and content words to be parsed as clausal or constituent reprise fragments, even though we have seen that some of these are very unlikely: it is the grounding process that will decide which analyses are preferable to others.

This approach certainly has its benefits (see section 6.3.5), but we must remember that the grammar as actually implemented here has a very narrow coverage, and a very small lexicon: it is designed only to support the prototype dialogue system and its domain while providing coverage of the CR phenomena above. If the grammar is extended to wider coverage, some degree of disambiguation might be required within the grammar itself to cut down on the number of possible alternatives that have to be processed.

Intonation

One source of possible disambiguating information that seems likely to be particularly important for CRs (and possibly for fragments in general) is intonation. Declarative bare answer fragments might be distinguishable from their interrogative equivalents, especially reprise CRs (see Srinivasan and Massaro, 2003), and even reprise CRs from other interrogatives (Grice et al., 1995). Also, the gap form seems to have a distinctive pitch contour (with a flat final "continuation" contour rather than a final rise or fall), and the clausal and constituent readings of reprise fragments may too (rise-fall-rise for clausal, and steady rise for constituent). Grammatical constraints could be designed to associate these contours with par-

---

[15]This is a very simplistic version of a standard robust parsing technique (see e.g. van Noord et al., 1999).

ticular readings (i.e. with particular values of MAX-QUD) as shown in AVM (319) for gaps.

$$
(319) \quad \begin{bmatrix} \text{INTON} & continuation \\ \text{CONT} & \boxed{1} \\ \text{HEAD-DTR} \mid \text{CONT} & \boxed{3} \\ \text{CTXT} & \begin{bmatrix} \text{MAX-QUD} & \boxed{1}\begin{bmatrix} ?\boxed{4}.\begin{bmatrix} utter\_consec(i,\boxed{3},\boxed{4}) \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}
$$

Such constraints are not applicable to the implementation described here (as it is text-based), but could be used in a version with a speech interface. However, given the variable accuracy of current pitch-tracking and frequency contour identification technology, and the often soft nature of intonational constraints, it seems likely that they would be most useful as weighted features in a probabilistic grammar rather than as strict logical constraints.

5.6   Summary

This chapter has shown how a HPSG grammar fragment can be defined which implements G&C's approach and extends it to cover the ontology of CRs developed in chapter 3, while including the semantic analysis of chapter 4. The various forms of CR are treated as follows:

Non-Reprise These CRs are analysed as standard sentences of the grammar, with the intended reading driven explicitly by the verb and its lexical semantics. Constituent and lexical readings are possible using utterance-anaphoric arguments.

Conventional Conventional CRs have their meaning specified in the lexicon, and can therefore be given any required reading directly.

Literal and *WH*-Substituted Reprises These follow G&C's analysis, with the connection to focus and MAX-QUD now made explicit, and the extensions of chapter 4 allowing many word and phrase types to be *wh*-substituted.

Reprise Fragments and Sluices Again, these follow G&C's analysis, with extension to other word and phrase types, and a modified representation of elliptical constructions via contextual abstraction.

Reprise Gaps and Fillers Gaps are analysed in the same way as elliptical reprise fragments, with a different QUD coercion mechanism assumed. The same can apply for fillers, but this has not been implemented.

Corrections These are not treated separately from other CRs, but given the analysis of standard reprise fragments as above.

In order to increase the modularity of the overall approach and dialogue system, and to keep future grammar modification to a minimum, the grammar has been kept with as wide

a coverage of CRs as possible: all the forms and readings that it can produce are allowed. Disambiguation (using the correlations observed in chapter 3) will then be performed by the grounding process, which is covered in the next chapter.

Chapter 6

The CLARIE System

Having examined the empirical nature of CRs, and proposed a suitable corresponding grammatical and semantic framework, we now turn to implementation. One of the main objectives of this work has been to produce a prototype dialogue system, CLARIE, capable of interpreting and producing the most important kinds of CR. This chapter describes this system, an information-state-based dialogue system incorporating the HPSG grammar of the previous chapter: it centres around a *grounding* process which both allows user CRs to be suitably interpreted and allows system CRs to be generated where necessary.

6.1   Introduction

The starting points of the implementation are the GoDiS dialogue system and the grammar of chapter 5. GoDiS and the TrindiKit provide the basic framework together with a starting point for information state (IS) and dialogue move engine (DME); in order to incorporate the required clarificational capabilities, however, significant changes are needed. The grammar, of course, is used as the basis for interpretation and generation.

It should be stressed that the system presented here is a prototype, intended as a proof of concept rather than as a fully-functioning dialogue system. As such, many elements that would be present in a full system – for example, a lexicon and grammar with suitable coverage, a realistic domain model and some inferential capability – are currently omitted for simplicity's sake. The prototype system is therefore very restricted in the dialogues it can sensibly handle, but does show how clarificational capability can be incorporated and what requirements this imposes on the IS and DME. It is also restricted to text-based input and output: the TrindiKit does allow speech-based input and output modules to be used, so this could be changed in future.[1]

---

[1]Plugging in a speech recogniser with a statistical language model would be relatively straightforward. However, many dialogue systems use grammar-based models to improve recognition rates, and interfacing such a model with the HPSG grammar used here, with its non-standard contextually dependent representation, might be

6.1.1   Aims and Requirements

The objective of this chapter is to show how a basic system can be built which can handle the most common forms of clarificational dialogue. In particular, the system should have the following two capabilities:

- User utterances which are problematic for the system (e.g. which contain unknown or ambiguous words or referents, or which contain noisy or incomprehensible parts) should be treated by appropriate clarification of the problematic part, followed by incorporation of any response.

- User utterances which are requesting clarification should be recognised correctly and responded to in an appropriate way.

In order to achieve these overall aims, the system will need to have a number of properties, resulting from the findings of the previous chapters. In particular, it will require a particular representation of utterances, and a particular treatment of the information state and DME update rules.

Linguistic Representation and Grammar

- The representation of utterances must include information at phonological, syntactic and semantic levels.

- This representation must have an appropriate semantic structure: it must be made contextually dependent, with words and certain phrases contributing elements which must be contextually identified during grounding.

- Both user and system utterances must share this representation, as both may be subject to clarification.

- Interpretation and disambiguation should be guided by the correlations so far described between word & phrase types and possible clarificational readings. This must also apply to the resolution of ellipsis.

- Sentences containing unknown or unrecognised words must be given an appropriate representation, so that (only) the problematic parts can be clarified.[2]

more of a challenge.

[2]Elements of the treatment of unknown words here have been previously published as (Purver, 2002).

Information State & DME

- The information state must include a record of utterances (rather than just e.g. dialogue moves). This may be of a limited length as discussed in chapter 3.

- A suitable grounding/interpretation process should apply contextual abstracts to the IS in order to fully specify utterances where possible, or result in suitable clarification if not.

- This process should also be guided by the correlations so far described, including the likely forms of answers to various question types.

- The construction of CRs, and of answers to user CRs, should follow the empirical correlations of chapter 3.

The next section 6.2 explains the immediate consequences of these requirements for the overall structure of the system and describes the changes to the GoDiS modules and the TrindiKit that this requires. Section 6.3 describes the grounding and integration process (the part of the DME which deals with user input, in particular interpreting user CRs as such, and deciding when to clarify input that cannot be grounded), and section 6.4 the selection and generation process (the part which deals with system output, in particular generating suitable CRs). Section 6.5 then summarises the overall approach, stepping through some sample dialogue extracts to illustrate how the various processes are integrated into the DME.

6.2   System Structure

CLARIE has the same overall system architecture as GoDiS, with various modules replaced to provide its new functionality. The keyword-based interpretation and generation modules are replaced with equivalents that use a full grammar (the HPSG grammar of chapter 5), and the DME update rules are replaced with a set which implement the approach to grounding and clarification which has been outlined so far.

The modular nature of GoDiS and the TrindiKit enables changes to particular aspects of the system to be made easily by replacing individual modules. In order to convert to an HPSG-based system, the interpretation and generation modules had to be replaced, and a new AVM interface resource module provided to allow other modules to use the resulting representations. To implement the new approach to grounding and clarification, changes to the DME modules (update and selection modules) were required.

6.2.1   Interpretation

The interpretation module of GoDiS, `interpret_simple`, uses a robust (although domain-specific) keyword/phrase-spotting method to turn an input string into a set of associated dia-

logue moves. Rather than a simple set of moves, the approach to grounding & clarification now proposed requires a full representation at semantic, syntactic and phonological levels, and a treatment of utterances as contextual abstracts. The system therefore uses the HPSG grammar already defined, which builds this kind of representation, and takes the output of the new module to be a set of *signs* (the multiple possible results produced by parsing the input string with a simple bottom-up left-to-right chart parser).

In keeping with the modular approach of the TrindiKit, the simple interpretation module is therefore replaced with a general grammar-based version `interpret_grammar` which can call any grammar implemented as a TrindiKit resource. As with GoDiS, the module provides one main predicate, `interpret/0`, which takes as input the IS variable `input`, a string corresponding to the latest user input. Whereas the GoDiS version produced as output the IS variable `latest_moves` (a set of dialogue moves), the CLARIE version now produces the variable `latest_utt` – a set of signs represented as AVMs. These signs encode dialogue moves as their content, together with all required phonological and syntactic information.

The grammar itself has already been described in chapter 5. The set of signs (in contextual abstract form) is passed to the grounding process to complete interpretation by fully instantiating content. This is specified within the DME update rules, and is described in section 6.3.

6.2.2   Generation

In GoDiS, generation was done via canned text – the domain-specific lexicon contained a set of pre-specified strings with their corresponding dialogue moves. The same approach would be possible here, but the fact that we have a grammar for interpretation allows a more general approach: the same grammar can be used for generation as interpretation. This has several advantages: it allows re-use of the same components rather than having to specify input and output patterns separately; it ensures that the system can interpret and generate the same set of possible sentences; and it ensures that both system and user utterances have the type and levels of representation that are required for treatment of clarification.[3]

CLARIE therefore uses a new generation module, `generate_grammar`, which uses a grammar resource in a directly parallel way to the interpretation module. A bottom-up chart generator has been implemented which uses the same grammar rules as the parser: as the grammar includes elliptical forms, this means elliptical utterances can be generated as well as interpreted. As with GoDiS, the new module takes as input the IS variable `next_move` (the required dialogue move), and updates the variable `output` with the resulting output string. It now also produces as an output the value of the variable `latest_utt`, the full sign representation corresponding to that dialogue move and string, as generated by the grammar, so that this can be used to update the information state with all possible utterance information

---

[3]For discussion of the advantages of interpretation/generation reversibility in general, see (Shieber, 1988; Erbach, 1991; Neumann, 1994), amongst others.

(just as for user input). It also has access to the variables `max_qud` and `sal_utt`, which allow generation of elliptical utterances to be limited to situations where they are desired.

This new generation module, together with the DME rules that decide on the next move to be generated, are described together in section 6.4.

6.2.3 AVM Representation

As the system uses feature structure representations of utterances and semantic objects throughout, a simple and efficient representation of AVMs is needed. While the TrindiKit's IS-handling capabilities do allow definition of feature structures (the IS itself is defined as an AVM with attributes of certain defined types) together with some functions for performing suitable operations on them, it does not allow some simple but extremely useful operations such as direct unification of features.

The grammar is implemented using ProFIT (Erbach, 1995), which not only allows AVMs to be handled as Prolog terms (thus allowing full Prolog unification), but also allows type inheritance, thus making it ideal for use with HPSG. However, it does require precompilation of the ProFIT source code into Prolog code. In order to allow the use of ProFIT for easy handling of AVMs throughout the system, while keeping the ProFIT precompilation to a minimum and thus allowing as many modules as possible to be written in pure Prolog, an interface between TrindiKit and ProFIT was added. Modules which use AVMs heavily (in this system, only the grammar) can be defined as ProFIT resources, written in ProFIT and compiled into Prolog on starting the system. Other modules handle AVMs via a new `avm` resource module: this gives a general interface to signs and other objects that the system has to manipulate (e.g. semantic objects such as propositions and questions), and hides the AVM structures themselves. This allows these modules (including the IS-handling modules that make up the DME) both to be written in Prolog and to be independent of the representation of semantic objects – only the AVM interface module need be changed if a different representation is desired. In theory, this approach allows grammar and representation to be changed on the fly (via a user-settable flag), just as can be done with domain and lexicon in the original GoDiS system.

6.2.4 Information State

GoDiS IS

The GoDiS IS is described fully in section 2.5.1, and is shown again here for reference as AVM (320):

$$
(320) \quad \begin{bmatrix}
\text{PRIVATE} & \begin{bmatrix}
\text{AGENDA} & \begin{bmatrix} stack(action) \end{bmatrix} \\
\text{PLAN} & \begin{bmatrix} stackset(action) \end{bmatrix} \\
\text{BEL} & \begin{bmatrix} set(proposition) \end{bmatrix} \\
\text{TMP} & \begin{bmatrix} shared \end{bmatrix}
\end{bmatrix} \\
\text{SHARED} & \begin{bmatrix}
\text{COM} & \begin{bmatrix} set(proposition) \end{bmatrix} \\
\text{QUD} & \begin{bmatrix} stack(question) \end{bmatrix} \\
\text{LU} & \begin{bmatrix}
\text{SPEAKER} & \begin{bmatrix} speaker \end{bmatrix} \\
\text{MOVES} & \begin{bmatrix} assocset(move,bool) \end{bmatrix}
\end{bmatrix} \\
\text{NIM} & \begin{bmatrix} stackset(move) \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

CLARIE IS

The IS used in the CLARIE system stays close to this overall structure. The private part of the IS remains very close, with only two changes (see AVM (321)). Firstly, due to the nature of the propositional structure that the system uses, a BG (short for BACKGROUND) record is required as well as the BEL record for storing private beliefs. BEL holds the propositional content (objects of type *proposition*), while BG holds objects of type *parameter* which contain information (e.g. names or other properties) about the indices which play roles in those propositions.

Secondly, CLARIE dispenses with the TMP record. As one of the main objectives is to model clarification by users, the grounding strategy adopted will not be of the optimistic nature that requires the kind of backtracking strategy that motivated this record. In future it could be replaced if an optimistic strategy is desired as an option, but it is left out here for

simplicity's sake.

$$
(321) \quad
\begin{bmatrix}
\text{PRIVATE} &
\begin{bmatrix}
\text{AGENDA} & \begin{bmatrix} \textit{stack(action)} \end{bmatrix} \\
\text{PLAN} & \begin{bmatrix} \textit{stackset(action)} \end{bmatrix} \\
\text{BEL} & \begin{bmatrix} \textit{set(proposition)} \end{bmatrix} \\
\text{BG} & \begin{bmatrix} \textit{set(parameter)} \end{bmatrix}
\end{bmatrix} \\
\text{SHARED} &
\begin{bmatrix}
\text{COM} & \begin{bmatrix} \textit{set(proposition)} \end{bmatrix} \\
\text{BG} & \begin{bmatrix} \textit{set(parameter)} \end{bmatrix} \\
\text{QUD} & \begin{bmatrix} \textit{stack(question)} \end{bmatrix} \\
\text{SAL-UTT} & \begin{bmatrix} \textit{stack(sign)} \end{bmatrix} \\
\text{UTT} & \begin{bmatrix} \textit{nstackset(4,sign)} \end{bmatrix} \\
\text{PENDING} & \begin{bmatrix} \textit{stack(set(sign))} \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

The shared part of the IS looks more different. The COM and QUD records are unchanged except that a BG record is added as before to hold parameter information about the indices that play roles in their propositions and questions. However, the storage of utterance information is different. The system represents utterances as *signs* so that all the levels of information are present; in fact, as *sets* of signs, as the grammar usually assigns more than one possible parse to an utterance (in fact always – see chapter 5). Signs include the move made by the utterance (as part of the sign's CONTENT) and the identity of the speaker (as part of the sign's contextually dependent C-PARAMS, fixed during grounding), so the individual SPEAKER and MOVES records are no longer required.

Following (Ginzburg and Cooper, 2004), a PENDING record is used for initial storage of these sets of signs before grounding. Once successfully grounded, utterances will be removed from PENDING; if grounding is impossible, they are left there while clarification takes place. This record is a *stack*: while it is possible to have nested clarification sequences (see sections 3.3 and 3.4), where more than one utterance must therefore be pending at once, we assume (as suggested by the results of section 3.4) that these sequences cannot be crossed, and that only the top (most recent) utterance therefore need be accessible – it must be grounded (or accommodated) and its clarification sequence closed off before the previous sequence can be returned to and the previous ungrounded utterance addressed.

Utterances are added to an UTT record, where they remain after grounding, providing a record of the utterances in the dialogue so far, which can be used both to identify sources for CRs and to provide the information required to answer them. This must allow more than one utterance to be stored, and must allow the operations of a set rather than just a stack (section 3.2 showed that CSS distances greater than one turn are common, in other words that it is possible to discuss any previous utterance, not just the immediately preceding one), but must also have a notion of linear order (questions such as *"What did you just say?"* or

"No, what did you say before that?" are always possible). The *stackset* datatype therefore seems suitable. However, given the large amount of information associated with each sign, and given considerations of memory, processor power and not least the ease of reading and understanding the IS, it is desirable to have the ability to limit the length of the record to prevent all sign information being kept forever. As shown in section 3.2, an utterance record of length 4 seems like a reasonable compromise (98% of CRs in the BNC sub-corpus were asking about utterances from 4 or fewer turns before). A new datatype was therefore added to those already provided with the TrindiKit, *nstackset*, which provides the operations of the *stackset* type but has a limited length n (in this system $n = 4$). It behaves as a FIFO (first in, first out) buffer, so contains the most recent n members only – as a new one is added, the oldest one drops out.

As will become clear in section 6.3, the PENDING record replaces GoDiS's NIM, as the grounding process removes utterances from it only when fully grounded with reference to the current IS. It also removes the need for the *assocset* type used for storing moves in GoDiS, as the setting of the associated flag now corresponds directly to moving from PENDING to UTT.[4]

The final change is the addition of a SAL-UTT stack: this is used to store certain grounded utterances for use in ellipsis resolution (along with the QUD stack). It performs a different function from the UTT record: while UTT maintains a record of utterances in the order in which they occurred in the dialogue, SAL-UTT maintains a record of utterances which are directly related to and have the same ordering as the members of QUD, which may not correspond to linear dialogue order (see sections 6.3 and 6.5 below for details).

6.2.5   Dialogue Management

With these changes, the DME can now follow the same general process as GoDiS: input and interpretation modules produce a representation of an incoming user utterance (in this case, a set of contextually dependent signs); an update module then applies rules to integrate this representation into the IS, determining what further effects it has and what the next resulting system action will be; a selection module determines a new move corresponding to that action; the generation and output modules then produce a suitable new system utterance and its representation; and the update module then once again integrates this into the IS together with its effects. The core part of the DME is therefore contained in the update module: it determines what effects utterances have on the common ground and how the system reacts to them.

There are two main departures from GoDiS as far as the update module is concerned: firstly, of course, the incorporation of the IS effects relating to clarificational dialogue; but secondly, and perhaps more significantly, in the conception of the grounding process. As

---

[4]IBiS, the successor of GoDiS, also manages without the *assocset* type by using a similar mechanism: NIM is used as PENDING is here.

utterances are now represented as contextual abstracts, they must be applied to the context (grounded) in order to be fully interpreted. The update rules therefore have to perform this task as well as determining the interaction of the fully specified utterance with the IS; as these two processes can affect one another they must be combined into a general update process, and this is described in section 6.3. The selection module also requires significant changes to incorporate clarificational dialogue, and is described in section 6.4.

6.3   Grounding and Integration

The system's ability to handle clarificational dialogue centres around the grounding process: the application of the contextually dependent utterance abstracts to the context, finding suitable values for each contextual parameter. It is the inability to ground a particular parameter in the current IS (or to ground it 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 (particularly elliptical forms) to be interpreted correctly. It also provides the method of disambiguation: grounding an ambiguous utterance involves finding a particular interpretation that can be grounded in a relevant way given the current context.

In CLARIE the grounding process is implemented in as simple a way as possible. No reasoning or inference is used; instead, a set of logical constraints and preferences that govern the process are defined as IS update rules. Prolog backtracking is then used to find a set of parameters such that all C-PARAMS are instantiated and all constraints are satisfied. The constraints are expressed as preconditions on particular rules, and express general requirements on the way parameters are instantiated: for example, to ensure that utterances are interpreted in such a way that their content is internally consistent and consistent with what is already known (where possible). The preferences are expressed in the ordering of the update rules, and ensure that utterances are made maximally relevant: for example, that an ambiguous utterance be taken as an answer to a question currently under discussion if it can function as such, and only taken as a CR if such an instantiation is not possible.

This should not be taken as an insistence that no reasoning or inference are necessary in a genuinely full treatment, or that they are not performed by humans – merely that the current implementation attempts to go as far as possible without them, to simplify the system and avoid the significant computational expense. There is no reason why the current treatment of grounding could not be combined with reasoning (e.g. the default logic used to reason about dialogue by (Asher and Lascarides, 2003; Schlangen, 2003)) to give a more complex yet more complete system.

6.3.1 The Basic Process

The grounding process is modelled on G&C's proposed utterance processing protocol. This is given (slightly simplified) in listing 6.1 below.

```
for utterance U in PENDING:
A:  if ( can find assignment f for U in IS )
    then ( add U to LATEST-MOVE,
           react to content of U,
           remove U from PENDING )
    else ( coerce MAX-QUD and SAL-UTT,
           goto A: ),
    else ( produce suitable clarification request )
```

Listing 6.1: G&C's Utterance Processing Protocol

The new utterance is first pushed onto the PENDING stack, then an attempt is made to apply its contextual abstract representation to the context (i.e. to ground it). If successful, it is removed from PENDING and reacted to; if not, a contextual coercion operation is used and grounding is re-attempted (i.e. now trying to ground it as a CR, with new coerced values of MAX-QUD and SAL-UTT). If all else fails, clarification ensues.

This structure is followed in general in the CLARIE update algorithm, with some differences, shown below in listing 6.2 (again, with some simplifications):

```
init,
repeat( integrate orelse
        coerce orelse
        accommodate orelse
        clarify ),
manage_agenda,
manage_plan,
manage_qud
```

Listing 6.2: Update algorithm

Here, before the grounding process begins, the `init` rules push the utterance (in fact, the set of possible ambiguous signs for that utterance, in their uninstantiated contextually abstracted form) onto both the PENDING stack and the UTT utterance record stack. Next, the `integrate` rules try to ground any one of those signs, given the IS in its current unchanged state. If a sign can be found for which a rule can instantiate the abstracted parameters in an acceptable way, the utterance (the set of signs) is popped from PENDING and any further effects associated with the type of move made are applied, integrating the fully instantiated move into the IS (adding commitments, downdating QUDs etc.). If not, the `coerce` rules try to ground the utterance by applying coercion operations (thus producing a modified MAX-QUD

and SAL-UTT); this is how CRs will be grounded. Again, if successful for any member of the set of signs, the utterance is popped from PENDING and its effects applied to the IS (as these moves will be CRs, they will raise new QUDs). If neither are successful, the `accommodate` rules then try to ground the utterance by accommodating planned but as yet not explicitly asked questions into the IS (as with GoDiS) – again, then applying effects and removing from PENDING. If all three of these fail, the utterance is left in PENDING, and an action to clarify some problematic feature is added to the agenda via the `clarify` rules. After this, some plan and QUD management rules apply to update the IS if necessary.

The `integrate`, `coerce` and `accommodate` processes therefore all perform both *grounding* (application of the utterance abstract to the context) and *integration* (update of the IS according to the move made). These two effects must be specified together in the same process as they are necessarily interdependent (see section 6.3.5 below). As these three sets of rules all end up with the utterance successfully grounded and removed from PENDING, we will refer to them together as *grounding rules*, or to the sets individually as *integration rules*, *coercion rules* or *accommodation rules* respectively. The `clarify` process on the other hand applies only to utterances that cannot be successfully grounded, and instead performs the job of selecting a new clarificatory action that must be taken by the system: we will refer to the rules that make it up as *clarification rules*.

Scaling Up   Note that this approach essentially treats of all the possible ambiguous signs as equally likely. This is a viable approach here, but of course the grammar and lexicon are small: scaling up to a wide-coverage grammar might generate very large numbers of possible signs for an utterance. In this case, a modified approach may be required, using probabilities or weights provided by the grammar to bias the grounding process, and considering most likely parses first, or perhaps only those above a certain likelihood threshold. Ideally, these parse probabilities could also be combined with contextually derived probabilities of a particular move being made (or a particular parameter being clarified). Either way, the basic process could remain the same – here only the simple version will be considered.

Grounding Rules

A typical grounding rule takes the form sketched out in listing 6.3, with preconditions checking for an instantiation that fulfils the general consistency constraints, and resulting effects that perform the required IS update. This is specified in the TrindiKit syntax as described in section 2.5.

This rule takes the first pending set of utterances, selects one member of that set, looks up its content and abstracted parameter set (via the `get_move` AVM interface condition), and then attempts to find a suitable instantiation of parameters (the `grounding` condition) given the current IS (specifically the values of MAX-QUD and SAL-UTT and the parameters

```
rule( groundingRule,
      [ fst( $/shared/pending, Set ),
        in( Set, Utt ),
        $avm :: get_move( Utt, lambda(Params,Move) ),
        fst( $/shared/qud, MQ ),
        fst( $/shared/sal_utt, SU ),
        $grounding :: grounding( Params, $/private/bg, $/shared/bg,
                                 MQ, SU, NewShared, Ungrounded ),
        $grounding :: consistent( Move, $/shared/com )
      ],
      [ pop( /shared/pending ),
        extend( /shared/bg, NewShared )
      ] ).
```

Listing 6.3: Grounding rule template

already in the current private and shared background), such that the result is consistent (the `consistent` conditions are met). Instantiation of parameters is performed via Prolog unification, so that if all conditions are met, the instantiated values remain. The `grounding` condition has access not only to the background parameter sets, as explicitly shown in listing 6.3, but also to the lexicon in order to identify known predicates (preventing all predicates in the lexicon having to be explicitly represented in the IS as part of the private background). These are returned as the `NewShared` argument. It also identifies which parameters (if any) cannot be grounded at all – rules usually check that this last `Ungrounded` argument is empty.

The effects of this typical rule remove the (now instantiated) utterance from the PENDING stack (it now remains only in the UTT record). The shared background is also extended where necessary to reflect the grounding of parameters to `NewShared` values which were previously only part of the lexicon or private background: now that they have been used in a public utterance they are considered explicitly shared.

Initialisation

Initialisation (the `init` process) is performed by a single rule `initialize`, shown in listing 6.4. The preconditions check the values of the IS variables `latest_speaker` and `latest_utt`, which have already been assigned by the interpretation process to the identity of the speaker, and the set of utterance abstracts output by the grammar, respectively. There are two further preconditions. The first checks that the two contextual parameters corresponding to speaker and addressee, which are present in all utterances, can be instantiated suitably; in this two-agent system this is trivial and will always succeed. The second similarly instantiates any existentially quantified parameters (members of QUANTS rather than C-PARAMS, which are also assigned to uninstantiated Prolog variables by the grammar) to atomic variables – this is merely a programming convenience which could in theory be performed by the grammar. The effects then assign the set of utterance abstracts (with these two parameters

instantiated, but no others) to the appropriate IS fields.

```
rule( initialize,
      [ $latest_speaker = Spkr,
        $latest_utt = USet,
        $grounding :: ground_participants( USet, Spkr ),
        $grounding :: instantiate_quants( USet )
      ],
      [ push( /shared/pending, USet ),
        push( /shared/utt, USet )
      ] ).
```

Listing 6.4: Grounding initialisation rule

Identification of speaker and addressee might not always be trivial, particularly in a multi-party dialogue system, where one of the jobs that must be performed in grounding will be precisely to establish who the speaker and the intended addressee are. In this case, instantiation of these parameters must be perfomed as part of the general grounding process (and will presumably require access to contextual information just as much as the grounding of other parameters – knowledge about beliefs and commitments, as well as features of the utterance itself, may affect who is taken to be the addressee). This is not necessary here, however, and treating these parameters separately helps simplify other grounding rules.

The next section 6.3.2 describes the `integrate` process, the rules which govern successful grounding and its after-effects; section 6.3.3 then describes the `coerce` process, the rules which perform the same purpose for CRs by using contextual coercion operations; then section 6.3.4 describes the `clarify` process, the rules which govern behaviour when grounding cannot take place. The `accommodate` process is not described here, but can be taken to follow GoDiS (see section 2.5) with modifications dictated by the new IS and utterance representation format; the same applies to the plan, agenda and QUD management rules.

6.3.2   Successful Grounding and Integration

This section describes the *integration* rules – those that successfully ground utterances and integrate them into the current IS, without requiring coercion. Given that there will be several possible representations produced by the grammar for each utterance, and possibly several ways of grounding each of these representations, the rules are ordered to express preferences over the possible interpretations that result. The preferences are currently as follows:

1. Interpret as answering a question which is under discussion.

2. Interpret as asking a question which is relevant to the current IS.

3. Interpret as a greeting, closing or thanking move.

4. Interpret as a CR.

The last option is the domain of the *coercion* rules described in the next section 6.3.3. It also contains many sub-options, of course, as CRs themselves are often ambiguous. These options are also ordered (according to the findings of chapter 3); this is described fully below.

Answers

The system first tries to treat the move made by the utterance as an answer to the question currently maximally under discussion (the first element of the /SHARED/QUD stack). The grounding condition must apply such that no parameters remain ungrounded, the consistency constraints hold, and the move made is an assertion of a proposition which answers the maximal question (listing 6.5).

```
rule( integrateAnswer,
      [ fst( $/shared/pending, USet ),
        in( USet, Utt ),
        $avm :: get_move( Utt, lambda(Params,Move) ),
        fst( $/shared/qud, MQ ),
        fst( $/shared/sal_utt, SU ),
        $grounding :: grounding( Params, $/private/bg, $/shared/bg,
                                 MQ, SU, NewShared, [] ),
        $grounding :: consistent( Move, $/shared/com )
        $avm :: move( Move, assert(P) ),
        $answerhood :: relevant_answer( MQ, P )
      ],
      [ pop( /shared/pending ),
        extend( /shared/bg, NewShared ),
        pop( /shared/qud ),
        pop( /shared/sal_utt ),
        add( /shared/com, P ),
        add( /shared/com, resolves(P,MQ) )
      ] ).
```

Listing 6.5: Grounding rule for answers

If these conditions hold, the effects of the rule then include the standard effects associated with successful grounding, together with the removal of the answered question from the stack of QUDs (together with its associated salient utterance from the stack of SAL-UTTs) and addition of two elements to the set of shared commitments: the answering proposition itself, and the fact that it resolved the question (see Ginzburg, forthcoming).

Answerhood is currently defined according to the simple method of (Macura, 2002): essentially, a proposition p answers a question $?\{\ldots\}.p$ (with certain restraints on quantification). The definition of answerhood is contained within its own TrindiKit resource module answerhood, so a more sophisticated approach could be easily substituted. One possible addition might be the classification of answers into those that *resolve* a question and those that

are merely *about* a question (see Ginzburg, 1995) – in this case separate update rules might
be required to treat the two classes of answers differently.

Questions

The next rule attempts to ground the utterance in such a way that its move asks a question
that is relevant to the current domain. Grounding and consistency conditions are applied as
before, such that the move becomes an `ask` move, asking a relevant question (defined as one
that influences the current plan). This check on relevance is important to allow irrelevant
moves to be clarified later, rather than successfully grounded (listing 6.6).

```
rule( integrateUsrAsk,
      [ fst( $/shared/pending, USet ),
        in( USet, Utt ),
        $avm :: speaker( Utt, usr ),
        $avm :: get_move( Utt, lambda(Params,Move) ),
        fst( $/shared/qud, MQ ),
        fst( $/shared/sal_utt, SU ),
        $grounding :: grounding( Params, $/private/bg, $/shared/bg,
                                 MQ, SU, NewShared, [] ),
        $grounding :: consistent( Move, $/shared/com )
        $avm :: move( Move, ask(Q) ),
        $answerhood :: influences( Q, $/private/plan ),
        $avm :: wh_utt( Utt, Wh )
      ],
      [ pop( /shared/pending ),
        extend( /shared/bg, NewShared ),
        push( /shared/qud, Q )
        push( /shared/sal_utt, Wh ),
        push( /private/agenda, respond(Q) )
      ] ).
```

Listing 6.6: Grounding rule for user questions

Here the effects push the new question onto the stack of QUDs (so that it is now maximal),
together with its associated *wh*-phrase (if any) onto SAL-UTT for ellipsis resolution, and push
a new action to respond to the question onto the agenda (so that this is the next action to be
processed). Note that this rule will not generally be able to apply for CRs, as they require
SAL-UTT coercion (see section 5.4).[5]

Questions asked by the system rather than the user must be treated differently (so that it
does not attempt to answer its own questions), and a different rule therefore applies to system
`ask` moves: no action to answer the question is added to the agenda, but instead the action to

---

[5]It is actually possible, though – non-reprise CRs (which do not require MAX-QUD coercion but only a suitable
value of SAL-UTT) which happen to be asking about the utterance which is currently already salient could meet
the conditions. But this doesn't matter – the effect will be to push the CR question onto QUD, which is exactly the
behaviour that would result from the CR rules anyway (see section 6.3.3 below). If this behaviour is not desired,
it can be prevented by checking that the question asked here is not a CR-type question.

raise the question (which caused it to be asked) is removed. There is also no need to check the question's relevance to the plan, as shown in listing 6.7 below.[6] As in GoDiS, most integration rules have different effects depending on the identity of the speaker: only the versions for the user will be shown from here on (see Larsson et al., 2000, for more detail on differences).

```
rule( integrateSysAsk,
      [ ...
        $avm :: speaker( Utt, sys ),
        ...
        $avm :: move( Move, ask(Q) ),
        $avm :: wh_utt( Utt, Wh ),
        fst( $/private/agenda, raise(Q) )
      ],
      [ pop( /shared/pending ),
        extend( /shared/bg, NewShared )
        push( /shared/qud, Q ),
        push( /shared/sal_utt, Wh ),
        pop( /private/agenda )
      ] ).
```

Listing 6.7: Grounding rule for system questions

A similar rule applies when the agenda action is `findout` rather than `raise`: in this case the action is not popped from the agenda on asking, but only when the question is answered (again, see Larsson et al., 2000).

Other Move Types

Rules for greetings, closings and thanks follow the general template, and also introduce agenda actions to e.g. return greetings if this has not already been done. This results in two variants of each rule, as shown for `integrateUsrGreet` below in listings 6.8 and 6.9.

This first version checks that the utterance can be grounded suitably, and that a previous system greeting can be found in the IS utterance record (and that this greeting can therefore be taken to be returning that original system greeting). The second version (listing 6.9) can apply only when no such original greeting can be found, and therefore adds a `greet` action to the agenda. Thanking and closing moves are treated similarly, although the resulting actions are to acknowledge and terminate the dialogue, respectively.

---

[6]Listing 6.7 is abbreviated in that it shows only those parts of the rule that differ from the previous version in listing 6.6. In order to make these rules more readable, they will be abbreviated from here on by omitting repeated standard sections where possible.

```
rule( integrateUsrGreet,
      [ ...
        $avm :: speaker( Utt, usr ),
        ...
        $avm :: move( Move, greet ),
        in( $/shared/utt, PrevUttSet ),
        in( PrevUttSet, PrevUtt ),
        $avm :: speaker( PrevUtt, sys ),
        $avm :: get_move( PrevUtt, lambda(_,PrevMove) ),
        $avm :: move( PrevMove, greet )
      ],
      [ pop( /shared/pending ),
        extend( /shared/bg, NewShared )
      ] ).
```

Listing 6.8: Grounding rule for return user greetings

```
rule( integrateUsrGreet,
      [ ...
        $avm :: speaker( Utt, usr ),
        ...
        $avm :: move( Move, greet ),
        not (
            in( $/shared/utt, PrevUttSet ) and
            in( PrevUttSet, PrevUtt ) and
            $avm :: speaker( PrevUtt, sys ) and
            $avm :: get_move( PrevUtt, lambda(_,PrevMove) ) and
            $avm :: move( PrevMove, greet )
            )
      ],
      [ pop( /shared/pending ),
        extend( /shared/bg, NewShared ),
        push( /private/agenda, greet )
      ] ).
```

Listing 6.9: Grounding rule for initial user greetings

6.3.3   Successful Grounding via Coercion

Grounding CRs requires SAL-UTT and possibly also MAX-QUD coercion. Additional ground-ing rules are therefore required to allow these CRs to be interpreted correctly by implement-ing G&C's *coercion operations*, which allow suitable values of MAX-QUD and SAL-UTT to be produced. The basic form of these rules therefore mirrors `integrateUsrAsk` except for this calculation of new contextual variables for use by the `grounding` condition, as shown in listing 6.10 below.

```
rule( integrateUsrCR,
      [ fst( $/shared/pending, USet ),
        in( USet, Utt ),
        $avm :: speaker( Utt, usr ),
        $avm :: get_move( Utt, lambda(Params,Move) ),
        in( $/shared/utt, SrcUSet ),
        in( SrcUSet, SrcUtt ),
        $avm :: speaker( SrcUtt, sys ),
        $avm :: constit( SrcUtt, Constit ),
        $grounding :: coercion_operation( SrcUtt, Constit, CQ, SU ),
        $grounding :: grounding( Params, $/private/bg, $/shared/bg,
                                 CQ, SU, NewShared, [] ),
        $grounding :: consistent( Move, $/shared/com ),
        $avm :: move( Move, ask(Q) )
      ],
      [ pop( /shared/pending ),
        extend( /shared/bg, NewShared ),
        push( /shared/qud, Q ),
        push( /private/agenda, respond(Q) )
      ] ).
```

Listing 6.10: Grounding rule for clarification questions

The difference in preconditions can be summarised as follows: a source utterance (spo-ken by the system – we assume no self-clarification by the user) is found in the utterance record and used in a coercion operation to form a new (focussed) CR question MAX-QUD and corresponding SAL-UTT value to ground the contextual parameters. The requirement that the question asked be relevant to the plan can of course be dropped – it has been effectively re-placed with the requirement that the question be a CR relevant to a recorded source utterance.

It is in the specification of these rules, their ordering, and in particular constraints on the nature of the source constituent, that the observations made in chapter 3 are used to disam-biguate the various elliptical CR forms and readings. The current protocol is as follows:

1. Coerce SAL-UTT only and interpret as a conventional or non-reprise CR.

2. Perform *parameter identification* and interpret as a constituent fragment reprise *if* the source is the first mention of a content phrase fragment.

3. Perform *parameter focussing* and interpret as a clausal reprise (sentence, fragment or sluice) *if* the source is a content phrase or number determiner.

4. Perform *gap identification* and interpret as a lexical reprise gap.

The first rule will allow conventional and non-reprise CRs, which have their readings specified by their syntax and/or semantics, to ask any CR question. The second ensures that constituent readings can only apply to reprise fragments (not sluices or sentences) and only when the source is a first mention and a content phrase. The third allows clausal readings of any form as long as the source is suitable, and the last then ensures that any remaining reprises (which must be of function words) are interpreted as gaps.

This seems to take care of all the most common form/reading combinations except one: it means that we are treating *wh*-substituted reprises as clausal rather than lexical, which goes against the findings of chapter 3. It would be entirely possible to define a rule and corresponding lexical coercion operation (see sections 5.4 and 3.2.5, and in particular AVM (86)) that does assign a lexical analysis to these CRs on the basis of their syntactic form – however, given that they are rare, and given that any system response generated will be the same as if it was a clausal CR (see section 6.4 below), this has not been implemented at present.

Conventional/Non-Reprise

The first rule (listing 6.11) applies only SAL-UTT coercion, providing as source any constituent of any utterance in the utterance record. Given the new TrindiKit definition of the *nstackset* type used by this record, the set membership predicate in will return the most recent utterances first (those nearest the top of the stack), thus ensuring that CRs are interpreted as being relevant to the most recent consistent utterance.

```
rule( integrateUsrCR,
      [ ...
        fst( $/shared/qud, MQ ),
        in( SrcUSet, SrcUtt ),
        $avm :: constit( SrcUtt, Constit ),
        $grounding :: grounding( Params, $/private/bg, $/shared/bg,
                                 MQ, Constit, NewShared, [] ),
        ...
```

Listing 6.11: Grounding rule for non-reprise CRs

As MAX-QUD is not being coerced in this rule, the standard value is used as already present in the IS. This would be required for any non-reprise CRs which are otherwise elliptical (although the current grammar does not include any).

Constituent Readings

These readings are constrained to apply only to content words or phrases (not function words) and to take as source an utterance of which no other occurrence can be found in the utterance record (therefore making the source the first mention), as in listing 6.12.

```
rule( integrateUsrCR,
      [ ...
        in( SrcUSet, SrcUtt ),
        $avm :: constit( SrcUtt, Constit ),
        $avm :: content_phrase( Constit ),
        not (
            in( $/shared/utt, PrevUSet ) and
            not ( PrevUSet == SrcUSet ) and
            in( PrevUSet, PrevUtt ) and
            $avm :: constit( PrevUtt, Constit )
            ),
        $grounding :: parameter_identification( SrcUtt, Constit, CQ, SU ),
        $grounding :: grounding( Params, $/private/bg, $/shared/bg,
                                 CQ, SU, NewShared, [] ),
        ...
```

Listing 6.12: Grounding rule for constituent CRs

Other constraints on particular forms are already expressed in the grammar (e.g. that constituent reprise fragments must be phonologically identical with their source) and will prevent the instantiation of the SAL-UTT and MAX-QUD parameters unless they are satisfied, causing the `grounding` condition to fail.

Clausal Readings

If the rule for constituent readings fails, the clausal rule shown in listing 6.13 is applied. This constrains the source to be either a content word or phrase, as before, or a number determiner (which seem to allow clausal readings – see chapter 3), but does not require it to be the first mention.

```
rule( integrateUsrCR,
      [ ...
        in( SrcUSet, SrcUtt ),
        $avm :: constit( SrcUtt, Constit ),
        ( $avm :: content_phrase( Constit ) or
          $avm :: number_determiner( Constit ),
        $grounding :: parameter_focussing( SrcUtt, Constit, CQ, SU ),
        $grounding :: grounding( Params, $/private/bg, $/shared/bg,
                                 CQ, SU, NewShared, [] ),
        ...
```

Listing 6.13: Grounding rule for clausal CRs

Lexical Readings

If the clausal and constituent rules fail, the utterance is treated as a reprise gap with a lexical reading (listing 6.14). This will therefore handle function words and other non-constituent strings, which cannot be standard reprise fragments.

```
rule( integrateUsrCR,
      [ ...
        in( SrcUSet, SrcUtt ),
        $avm :: constit( SrcUtt, Constit ),
        $grounding :: gap_identification( SrcUtt, Constit, CQ, SU ),
        $grounding :: grounding( Params, $/private/bg, $/shared/bg,
                                 CQ, SU, NewShared, [] ),
        ...
```

Listing 6.14: Grounding rule for reprise gaps

Over-Answering Polar Questions

As noted in chapter 3, there may be some cases where yes/no questions should be answered not only directly (with a yes/no answer), but also given a secondary supplementary answer. Hockey et al. (1997) suggest that this is more likely to be required for negative answers, and indeed this seems intuitively to be the case for CRs – in the invented example (322) below, the bare answer *"No"* seems most unhelpful, although in the affirmative equivalent example (323), bare answers seem much more acceptable:

(322)
>| A: | Did Bo leave? |
>| B: | JO? |
>| A: | No. / No, BO. |

(323)
>| A: | Did Bo leave? |
>| B: | BO? |
>| A: | Yes. / Uh-huh. / Yes, Bo. |

Given the range of CRs we are dealing with here, the only ones for which this is relevant are polar clausal questions, i.e. reprise sentences and fragments. Gaps, constituent fragments, conventional CRs and the *wh*-reprises all ask *wh*-questions rather than yes/no questions.

Wahlster et al. (1983) see this kind of supplementary answer as implicitly answering a further question: rather than just answering (negatively) the polar question *"Did you ask whether Jo left?"*, an answer to the related *wh*-question *"Who$_i$ did you ask whether i left?"*. Now, given the analysis via parameter focussing, this question is precisely the MAX-QUD. This offers a simple way of ensuring that such answers are given: an action to respond to the MAX-QUD question can be explicitly added to the agenda.

```
rule( integrateUsrCR,
      [ ...
        in( SrcUSet, SrcUtt ),
        $avm :: constit( SrcUtt, Constit ),
        ( $avm :: content_phrase( Constit ) or
          $avm :: number_determiner( Constit ),
        $grounding :: parameter_focussing( SrcUtt, Constit, CQ, SU ),
        $grounding :: grounding( Params, $/private/bg, $/shared/bg,
                                 CQ, SU, NewShared, [] ),
        $avm :: move( Move, ask(Q) ),
        $avm :: question( Q, [], _P ),
        ...
      ],
      [ ...
        push( /private/agenda, respond(Q,CQ) )
      ] ).
```

Listing 6.15: Over-answering clausal CRs

As shown in listing 6.15, then, a version of the clausal CR grounding rule can be formulated which checks that the question asked is polar (has an empty PARAMS set), and adds the coerced MAX-QUD question to the resulting `respond` action.[7] The version shown here will do this for any polar clausal CR – a similar more complex version could be formulated if desired which only results in this action if the question will be answered negatively, but it seems simpler to leave this decision until the answering proposition is determined in the selection module – see section 6.4 below.

6.3.4   Unsuccessful Grounding and Clarification

If none of the grounding rules can apply, the DME moves to the *clarification* rules, which are described in this section. These rules do not remove the utterance from the PENDING stack (as grounding has failed) but instead cause a clarification question to be asked: an action is added to the agenda which causes the system to generate this question as its next task. The utterance is left on the pending stack and will only be removed if an answer to the clarification question (or other information provided in some other way) subsequently allows the problematic parameter(s) to be instantiated and the utterance fully grounded. Note that these rules will apply to any utterance which cannot be grounded by the integration or coercion rules, including CRs for which a suitable source utterance cannot be found – so CRs-of-CRs, which we saw in chapter 3 can exist, can certainly be generated if needed.

These rules do not specify exactly which CR form is to be used or even which CR reading it will have – they just add an agenda action to clarify the utterance and/or problematic parameter. The choice of form and reading is left to the selection and generation modules,

---

[7]There is actually a further complication here – in order to license elliptical answers to both questions later, both questions must be pushed onto the QUD stack, rather than just the explicitly asked question.

described below in section 6.4.

Uninterpretable Utterances

The simplest case is one in which the parser could not assign a representation to the utterance. In this case, grounding cannot proceed, and instead a CR concerning the whole utterance must be asked instead (say, a conventional form indicating complete incomprehension). This is performed by the rule `clarifyUnknownUtterance` as shown in listing 6.16.

```
rule( clarifyUnknownUtterance,
      [ fst( $/shared/pending, USet ),
        not (
             in( USet, Utt ) and
             $avm :: get_move( Utt, _ )
             )
      ],
      [ push( /private/agenda, clarify(USet) )
      ] ).
```

Listing 6.16: Clarification rule for uninterpretable utterances

The preconditions check directly that no representation has been assigned by the interpretation process (`get_move/2` fails). The effect of the rule is to add an action to the agenda which will cause a CR question to be asked on the next system turn.

In this implementation, as there is no speech interface, such cases will always be due to grammar coverage problems. There is therefore no need to draw a distinction between cases where the words could not be perceived (due to e.g. noisy environment) and cases where the words were perceived but the grammar could still not parse the string. Indeed, as many spoken dialogue systems use grammar-based speech recognisers, this distinction may not be necessary even in these cases. If this distinction is required, separate rules which cause distinct questions to be raised could be used.

Unknown Parameters

The parameter set is first checked for parameters that cannot be instantiated in any way (corresponding to unknown names or words that are outside the lexicon), or are ambiguous between more than one possible referent. This is performed by the grounding rule `clarifyUnknownParameter`, shown in listing 6.17, which leads to a CR concerning the intended content of the problematic parameter.

The preconditions of this rule check that the last argument of the `grounding` condition, the ungrounded parameters, is *not* empty, but contains at least one parameter that cannot be grounded at all, and then check for the constituent of the utterance that contributed this

```
rule( clarifyUnknownParameter,
      [ fst( $/shared/pending, USet ),
        in( USet, Utt ),
        $avm :: get_move( Utt, lambda(Params,Move) ),
        fst( $/shared/qud, MQ ),
        fst( $/shared/sal_utt, SU ),
        $grounding :: grounding( Params, $/private/bg, $/shared/bg,
                                 MQ, SU, NewShared, [unknown(P) | _] ),
        $avm :: constit( Utt, Constit ),
        $avm :: content( Constit, P )
      ],
      [ push( /private/agenda, clarify( Utt, Constit, unknown(P) ) )
      ] ).
```

Listing 6.17: Clarification rule for unknown parameters

parameter. The effect is again to add the corresponding action (including information not only about the whole utterance, but the problematic constituent and parameter). The same principle applies for ambiguous parameters, as shown in listing 6.18, although a slightly different action is added to the agenda, in order to distinguish the cause so that a different CR form or reading can be generated if desired.

```
rule( clarifyUnknownParameter,
      [ fst( $/shared/pending, USet ),
        in( USet, Utt ),
        $avm :: get_move( Utt, lambda(Params,Move) ),
        fst( $/shared/qud, MQ ),
        fst( $/shared/sal_utt, SU ),
        $grounding :: grounding( Params, $/private/bg, $/shared/bg,
                                 MQ, SU, NewShared, [ambig(P,Alts) | _] ),
        $avm :: constit( Utt, Constit ),
        $avm :: content( Constit, P )
      ],
      [ push( /private/agenda, clarify( Utt, Constit, ambig(P,Alts) ) )
      ] ).
```

Listing 6.18: Clarification rule for ambiguous parameters

Inconsistent Parameters

In cases where all parameters can be grounded, the resulting instantiated utterance may still not be internally consistent, or may not be consistent with the common ground. The first rule that can apply to such cases is `clarifyInconsistentParameter`, shown in listing 6.19 below, which applies in cases which are only externally inconsistent, and when one parameter can be found that appears to be the cause of this inconsistency: that if it could be given another value available in the background, the resulting instantiated move would be

consistent.

```
rule( clarifyInconsistentParameter,
   [ fst( $/shared/pending, USet ),
     in( USet, Utt ),
     $avm :: get_move( Utt, lambda(Params,Move) ),
     fst( $/shared/qud, MQ ),
     fst( $/shared/sal_utt, SU ),
     $grounding :: grounding( Params, $/private/bg, $/shared/bg,
                              MQ, SU, NewShared, [] ),
     $grounding :: consistent( Move, _AnyCom ),
     $grounding :: inconsistent( Move, $/shared/com, $/shared/bg, Prop, P ),
     $avm :: constit( Utt, Constit ),
     $avm :: content( Constit, P )
   ],
   [ push( /private/agenda, clarify( Utt, Constit, inconsistent(P) ) ),
     del( /shared/com, Prop )
   ] ).
```

Listing 6.19: Clarification rule for inconsistent parameters

An example of such a case might be where the system believes the user wants to go Paris (and such a proposition is present in /SHARED/COM), and the only way that the latest utterance can be grounded is as an assertion that the user wants to go to London. In such a case, the parameter $x : name(x, london)$ is taken as the problematic parameter. The agenda action will eventually end up causing a clausal clarification question to be produced (corresponding to the "surprise" clausal question which might take the form *"London?"*).

The identification of the problematic parameter is done (in a simplistic way) by the `inconsistent/5` condition: if a parameter can be substituted by one from the shared background such that the resulting move is now consistent with the common ground, it is taken to be the problematic parameter and an action to clarify it is added. The effects of the rule also remove the conflicting proposition from COM now that there is doubt about it – any answer to the CR will cause a replacement proposition to be added instead (see section 6.5.3).[8]

Inconsistent Moves

If this rule fails (if no particular parameter can be identified as the cause of the problem) then a more general rule `clarifyInconsistentMove` applies (see listing 6.20) which will produce a clarification question which queries the whole move; again, this will be a polar clausal question which asks whether the move made was really as has been understood by the system. Here the preconditions just have to check that the consistency constraints fail.

---

[8]There is no doubt that deleting a single proposition from a set is a simplistic approach to belief revision, but given the very simple domain used here it is enough for the current purposes.

```
rule( clarifyInconsistentMove,
      [ fst( $/shared/pending, USet ),
        in( USet, Utt ),
        $avm :: get_move( Utt, lambda(Params,Move) ),
        fst( $/shared/qud, MQ ),
        fst( $/shared/sal_utt, SU ),
        $grounding :: grounding( Params, $/private/bg, $/shared/bg,
                                 MQ, SU, NewShared, [] ),
        not $grounding :: consistent( Move, $/shared/com )
      ],
      [ push( /private/agenda, clarify( Utt, Move, inconsistent ) )
      ] ).
```

Listing 6.20: Clarification rule for inconsistent moves

Irrelevant Moves

The final case in which grounding is considered unsuccessful is when an utterance cannot be
grounded in such a way as to be relevant to the current QUD or plan – in other words, where
all attempts to ground in a relevant way as in sections 6.3.2 and 6.3.3 have failed, but specific
interpretation problems such as already set out in this section have also not been identified.
This case can therefore be handled by a "failsafe" rule (listing 6.21) which catches all cases
which have not been dealt with by a rule so far.

```
rule( clarifyIrrelevantMove,
      [ fst( $/shared/pending, USet ),
        in( USet, Utt ),
        $avm :: get_move( Utt, lambda(Params,Move) ),
        fst( $/shared/qud, MQ ),
        fst( $/shared/sal_utt, SU ),
        $grounding :: grounding( Params, $/private/bg, $/shared/bg,
                                 MQ, SU, NewShared, [] ),
        $grounding :: consistent( Move, $/shared/com )
      ],
      [ push( /private/agenda, clarify( Utt, Move, irrelevant ) )
      ] ).
```

Listing 6.21: Clarification rule for irrelevant moves

No extra preconditions therefore need to be specified for this rule: they check that the
utterance can be grounded in a consistent way before the question is formed (to ensure that
the question can convey the interpretation that the system is, possibly mistakenly, giving to
the utterance), but it is not removed from the pending stack. In the current implementation,
the utterances to which this rule will apply includes those whose MAX-QUD or SAL-UTT
parameters cannot be grounded (and whose relevance therefore cannot be established), but is
not restricted to them – explicit non-elliptical questions or assertions which are not relevant

to the current plan or context will also be included.

6.3.5   Grounding vs. Integration

The grounding rules described in sections 6.3.2 and 6.3.3 perform more than one function: they ground an utterance, instantiating its contextual parameters, and then they apply its IS update effects, e.g. adding or removing QUDs. It might seem cleaner (and more modular) to separate the processes: one set of rules to perform grounding (presumably including consistency checking), and one set of integration rules to apply the resulting effects.

However, there are two reasons not to do this. Firstly, defining the grounding process in terms of all-in-one IS update rules means it becomes an integral part of the IS update process. This means that not only is the entire IS, with all the information contained therein, available during grounding, but so are the utterance's potential effects. This can be vital when trying to disambiguate the different possible interpretations of an utterance, and the different possible ways of grounding it. Grounding can be dependent on properties of the utterance itself (syntax and semantics), on pragmatic contextual information in the IS (previous utterances, private beliefs and the commitments so far built up in the common ground), *and* on the move that the utterance will make and the effects it will have. As observed by e.g. Schlangen and Lascarides (2002); Schlangen et al. (2003), this can be desirable for the correct disambiguation and resolution of fragments and other types of underspecified utterances. Information can flow in two directions: determining reference of parameters can establish which of the interpretations or moves are possible, and the possible moves and their effects can determine the reference of parameters. A fragment *"Bo"* may be a direct question, a declarative answer, or one of various types of CR: determination of which move is being made and of what the intended referent of *Bo* is (utterance-anaphoric or otherwise) are dependent on each other and best done together.

Secondly, separating grounding from integration would make the clarification of irrelevant moves described in section 6.3.4 extremely difficult. For this kind of clarification to be possible, utterances must not be considered grounded unless the move they make is relevant to the current plan or discourse. This can only be determined by checking whether their move can be integrated into the current IS – in other words, taking integration (or rather its failure) into account during grounding, thus making two separate sets of rules entirely dependent on one another (and thus making their separation pointless).

6.3.6   The `grounding` Resource

The process of instantiating contextual parameters is performed by a new TrindiKit resource module, `grounding` – this provides the grounding and consistency conditions used by all the rules described so far, as well as the contextual coercion operations. This section gives a brief description of this module and how the grounding process is actually defined.

Overview

The main interface predicate provided is `grounding/7`, which describes a relation between the following arguments:

1. A set of contextual parameters to be instantiated.

2. A set of known parameters from the IS private background.

3. A set of known parameters from the IS shared background.

4. The current value of MAX-QUD.

5. The current value of SAL-UTT.

6. A set of newly instantiated parameters to be added to the IS shared background.

7. A set of parameters which cannot be satisfactorily instantiated.

Those parameters in the original set (argument 1) that can be successfully uniquely instantiated to values given by context (arguments 2–5) have their INDEX values unified with the corresponding contextually provided referents. Those that cannot are made members of the "problematic" set (argument 7) which is used by the grounding rules of section 6.3.4 above as the source of clarification questions. Argument 6 is a set of parameters that can be uniquely instantiated, but only to values provided by outside resources rather than the IS itself (primarily the lexicon as described in section 6.3.1 above), or only to values in the private background, and must therefore be added to the shared part of the IS. The basic process is illustrated by the top-level predicate definition in listing 6.22.[9]

Parameter Instantiation

Parameters can be instantiated in the following ways:

1. Its INDEX value can be unified with a (set of) referent(s) taken from parameters in the background set which uniquely satisfy the specification of the RESTR set.

2. Its INDEX value can be unified with a referent provided by the domain, which uniquely satisfies the specification of the RESTR set (which in this case might be a name or other description).

3. Its INDEX value can be unified with a relation provided by the lexicon, which uniquely satisfies the specification of the RESTR set (which in this case will be a relation name).

---

[9]The version shown here is simplified in that it treats private and shared background together. The full version follows the same schema, but must also add parameters found in only the private background to the `New` set for introduction in to the shared part of the IS. Details of MAX-QUD and SAL-UTT parameter handling are also not shown – they must be treated separately to the general `unknown(P)` case.

```
% base case
grounding( [], _BG, _MQ, _SU, [], [] ).
% successfully ground to unique referent in background
grounding( [P | Params], BG, MQ, SU, New, Ungrounded ) :-
    unique_ref_in( P, BG ),
    !,
    grounding( Params, BG, MQ, SU, New, Ungrounded ).
% successfully ground to unique referent from lexicon/domain
grounding( [P | Params], BG, MQ, SU, [P | New], Ungrounded ) :-
    unique_ref( P ),
    !,
    grounding( Params, [P | BG], MQ, SU, New, Ungrounded ).
% unsuccessful because ambiguous
grounding( [P | Params], BG, MQ, SU, New, [ambig(P,A) | Ungrounded] ) :-
    ambig_ref_in( P, BG, A ),
    !,
    grounding( Params, BG, MQ, SU, New, Ungrounded ).
% ground max-qud
grounding( [P | Params], BG, MQ, SU, New, Ungrounded ) :-
    index( P, MQ ),
    !,
    grounding( Params, BG, MQ, SU, New, Ungrounded ).
% ground sal-utt
grounding( [P | Params], BG, MQ, SU, New, Ungrounded ) :-
    index( P, SU ),
    !,
    grounding( Params, BG, MQ, SU, New, Ungrounded ).
% otherwise unsuccessful
grounding( [P | Params], BG, MQ, SU, New, [unknown(P) | Ungrounded] ) :-
    grounding( Params, BG, MQ, SU, New, Ungrounded ).
```

Listing 6.22: Grounding schema

4. Its INDEX value can be unified with the MAX-QUD value.

5. Its INDEX value can be unified with the SAL-UTT value.

This definition effectively states that values available in context will take precedence over values provided by the domain or lexicon but *not* present in the IS context itself (i.e. that have not been raised so far in the dialogue). For example, a definite description "the destination" could be successfully resolved if exactly one destination has been discussed and previously added to context, even if the domain defines several possible destinations; similarly an ambiguous noun or verb which has been previously used (and successfully grounded, possibly after clarification) in one particular sense will be interpreted in this way again. This seems to fit with common approaches to reference resolution in which referents are only required to be unique in the current situation or sphere of attention (see e.g. Poesio, 1993) and also with psycholinguistic observations on the alignment of words with particular meanings between conversational participants (see Pickering and Garrod, 2004).

Note that as long as the lexicon uniquely describes the relations denoted by function words, clarifications of such words will not arise, so there is no need to specifically prevent them from being generated or to define grounding behaviour which differs specifically between content and function words. In fact, given a rich enough lexicon and a sentence in which the constraints on argument roles allowed more than one possible interpretation of a function word, grounding failure and subsequent CRs would be possible, and this seems like the correct behaviour.

Consistency Checking

The consistency check is currently only a simplistic version and takes the following form. A proposition p is taken to be *inconsistent* with a set of propositions (or facts) if:

1. The set contains $\neg p$.

2. The set contains p' where both p and p' contain the same relation R, differing in the indices that fill argument roles only in the nth argument, where R is defined in lexicon or domain to be unique in the nth argument. This allows e.g. the propositions that the user wants to go to Paris and that the user wants to go to London to be considered inconsistent in the travel agent domain.

Any proposition which is not inconsistent is then taken to be consistent. This definition could of course be much improved (particularly if logical inference were made available) in a full-scale system. Note that the second condition above is used as the basis for finding an inconsistent *parameter* (see section 6.3.4 above) – defined as the parameter associated with the conflicting nth argument position.

One point worth noticing here is that the consistency check is performed separately to the grounding (parameter instantiation) process. This prevents consistency being used as a constraint on parameter instantiation; situations where an ambiguous parameter could be instantiated in more than one way, but only one is consistent with context, could potentially be helped by a different approach where the two processes are combined. This is perfectly feasible (although it would require quite a different implementation of the instantiation process – see the sketch in listing 6.23 below), but especially given the simplicity of the current consistency check, a safer method seems to be to prevent grounding in such cases (so that they are clarified).

```
findall( RefSet, (
                possible_grounding( Params, SBg, PBg, MQ, SU, RefSet ),
                consistent( Com, Bel, Move ),
                ), RefSets ),
length( RefSets, 1 ).
```

Listing 6.23: Grounding schema with consistency check

Coercion Operations

The contextual coercion operations are defined straightforwardly in terms of properties of the source utterance. The versions given in chapter 5 can be implemented directly, as shown in listing 6.24 below for the parameter focussing operation:

```
parameter_focussing( Utt, Constit, MaxQud, SalUtt ) :-
    c_params( Utt, Params ),
    member( Param, Params ),
    content( Constit, Param ),
    content( Utt, Move ),
    constit( Utt, SalUtt ),
    constit( SalUtt, Constit ),
    question( MaxQud, [Param], Move ).
```

Listing 6.24: Parameter focussing coercion operation

The parameter identification and gap identification operations similarly follow the versions of chapter 5.

6.4   Selection and Generation

The previous sections have described how the interpretation process works, by parsing utterances with a grammar to produce contextually-dependent signs, then grounding these in the IS context via the DME update rules. Amongst other things, the effects of these update rules are to add new actions to the agenda, causing the system to raise new questions, or answer questions currently under discussion. These actions must then be turned first into dialogue moves (a task performed by DME selection rules, which are the main subject of this section) and then into surface strings which can be output to the user (performed by the generation module).

The basic non-clarificatory selection rules are relatively simple and are taken directly from GoDiS, with adjustments for the AVM representation used. Rules are specified for each type of agenda action, creating the required dialogue move from the action and its associated content. These are described in section 6.4.1. Selection rules for answering user CRs are also relatively simple, but differ in that they use the utterance record as a basis for producing an answer – see section 6.4.2. Rules for producing system CRs (section 6.4.3) behave similarly, but must also decide between the various CR *forms* that can be used to express the same *reading*; in particular, the choice between reprise and non-reprise forms, and between elliptical and non-elliptical. Section 6.4.4 then briefly describes the move-to-string generation process.

6.4.1   The Basic Process

The two main non-clarificatory actions that need to be handled are those for asking and answering questions, which correspond to `raise` or `findout` and `respond` actions respectively. These look very similar to the GoDiS equivalents, with the additional use of the `avm` AVM interface to create dialogue moves from the corresponding move type and semantic content. As shown in listings 6.25 and 6.26, the move is created directly and the `next_moves` variable is set in order to pass the move to the generation module. For questions, the move is of course to ask the question (listing 6.25):

```
rule( selectAsk,
      [ fst( $/private/agenda, findout(Q) ) or
        fst( $/private/agenda, raise(Q) ),
        $avm :: move( M, ask(Q) )
      ],
      [ set( next_moves, set([M]) )
      ] ).
```

Listing 6.25: Selection rule for questions

For answers, an answering proposition must be found and the new move is to assert this proposition (listing 6.26). The proposition is simply found in context (in the private beliefs or shared commitments) – propositions concerning e.g. ticket price which are not present at the start of the dialogue must be added to these beliefs as new information is gathered.

```
rule( selectAnswer,
      [ fst( $/private/agenda, respond(Q) ),
        in( $/private/bel, P ) or in( $/shared/com, P ),
        $answerhood :: relevant_answer( Q, P ),
        $avm :: move( M, assert(P) )
      ],
      [ set( next_moves, set([M]) )
      ] ).
```

Listing 6.26: Selection rule for answers

Moves such as greetings and closings, where the action type corresponds directly to the move type, and content consists entirely of the move type, are handled by a single generalised rule as shown in listing 6.27 below: the move is created directly and the `next_moves` variable set.

6.4.2   Answering User CRs

The selection rules for answering user CRs are in principle very similar to the `selectAnswer` rule above, in that the top agenda action is to respond to a particular question, and the next

```
rule( selectOther,
      [ fst( $/private/agenda, Action ),
        $avm :: move( M, Action )
      ],
      [ set( next_moves, set([M]) )
      ] ).
```

Listing 6.27: Selection rule for greetings etc.

selected move is an assertion which functions as an answer to that question. With CR cases, though, the answering proposition cannot be found directly in COM or BEL as a shared or private belief, but must be generated on the basis of the IS itself, specifically the properties of the source utterance being queried.

The rules must therefore take the following form: the source utterance (produced by the speaker – again, we are assuming no self-clarification by the user) must be found in the UTT record, and its properties can directly provide a proposition which answers the question. In almost all cases (all CRs except certain *wh*-question non-reprise CRs like *"What did you say?"*), the source utterance has already been identified during grounding and unified with the CR's SAL-UTT parameter, so finding the source utterance now is merely a trivial question of selecting the member of UTT which can unify with the CR question in a suitable way. For the few non-reprise questions for which this is not the case, the most recent suitable member of UTT will be taken, as the *nstackset* type returns members nearest the top first – this is of course what we want in cases like *"What did you say?"*.

Clausal Answers Clausal questions ask about the move made by the source utterance, and thus can be answered by a proposition concerning that move, which can be taken directly from the content of the source utterance (which is grounded in the system's IS representation and therefore has all referents fixed). All we have to do is find a source utterance whose content (a move) provides an answer, as in listing 6.28 below.

```
rule( selectAnswerCR,
      [ fst( $/private/agenda, respond(Q) ),
        $avm :: question( Q, _Params, P ),
        $avm :: illoc_rel( P ),
        in( $/shared/utt, USet ),
        in( USet, Utt ),
        $avm :: speaker( Utt, sys ),
        $avm :: content( Utt, P ),
        $answerhood :: relevant_answer( Q, P ),
        $avm :: move( M, assert(P) )
      ],
      [ set( next_moves, set([M]) )
      ] ).
```

Listing 6.28: Selection rule for clausal CR answers

Constituent Answers   For constituent questions, an antecedent utterance must be found with a constituent whose content fills the role asked about by the CR, as in listing 6.29.

```
rule( selectAnswerCR,
      [ fst( $/private/agenda, respond(Q) ),
        $avm :: question( Q, [Param], P ),
        $avm :: spkr_meaning_rel( P, sys, Constit, Param ),
        in( $/shared/utt, USet ),
        in( USet, Utt ),
        $avm :: speaker( Utt, sys ),
        $avm :: constit( Utt, Constit ),
        $avm :: content( Constit, Param ),
        $answerhood :: relevant_answer( Q, P ),
        $avm :: move( M, assert(P) )
      ],
      [ set( next_moves, set([M]) )
      ] ).
```

Listing 6.29: Selection rule for constituent CR answers

Alternative Descriptions   When answering a constituent question about the meaning of a word or the referent of a proper name, using the same word in the answer is unlikely to be very useful (e.g. answering *"Who do you mean 'Bo'?"* with *"Bo"*). What is required is an alternative description of the queried referent (*"My brother"*). If one is available as a parameter in the shared background (or in the domain or lexicon), it can be used instead of the original content of the source utterance. All the rule needs to do is find a parameter in the background which has the same referent as the original source constituent (but is not actually identical to the parameter associated with that constituent, i.e. the original description), as shown in listing 6.30.

If this fails (if no alternative description can be found) then the basic version of the rule in listing 6.29 is used. More sophisticated methods of generating the alternative description are of course possible and would certainly be required in a full-scale system. One obvious move might be the use of a database such as WordNet (Fellbaum, 1998) to generate alternative descriptions.

```
rule( selectAnswerCR,
      [ fst( $/private/agenda, respond(Q) ),
        $avm :: question( Q, [Param], P ),
        $avm :: spkr_meaning_rel( P, sys, Constit, Param ),
        in( $/shared/utt, USet ),
        in( USet, Utt ),
        $avm :: speaker( Utt, sys ),
        $avm :: constit( Utt, Constit ),
        $avm :: content( Constit, Param1 ),
        in( $/shared/bg, Param ),
        not ( Param = Param1 ),
        $avm :: co_referent( Param1, Param ),
        $answerhood :: relevant_answer( Q, P ),
        $avm :: move( M, assert(P) )
      ],
      [ set( next_moves, set([M]) )
      ] ).
```

Listing 6.30: Selection rule for constituent CR answers (alternative)

Lexical Answers   A similar version can be formulated for standard lexical questions, as in listing 6.31: here, a source utterance (or sub-constituent) must be found which provides an answer directly (i.e. by its existence in the UTT record).[10]

```
rule( selectAnswerCR,
      [ fst( $/private/agenda, respond(Q) ),
        $avm :: question( Q, _Params, P ),
        $avm :: utter_rel( P, sys, Constit ),
        in( $/shared/utt, USet ),
        in( USet, Utt ),
        $avm :: speaker( Utt, sys ),
        $avm :: constit( Utt, Constit ),
        $answerhood :: relevant_answer( Q, P ),
        $avm :: move( M, assert(P) )
      ],
      [ set( next_moves, set([M]) )
      ] ).
```

Listing 6.31: Selection rule for lexical CR answers

Lexical Gap Answers   For gap questions, again the overall approach is similar; in this case, a source utterance must be found with two consecutive constituents which can fill the roles asked about in the question (see listing 6.32).

---

[10]In fact, a full treatment must allow negative and affirmative answers to polar lexical questions (*"Did you say 'Paris'?"*) as well as the *wh*-question shown here. This can of course follow the treatment of negative and affirmative polar clausal questions below.

```
rule( selectAnswerCR,
      [ fst( $/private/agenda, respond(Q) ),
        $avm :: question( Q, [Constit2], P ),
        $avm :: utter_consec_rel( P, sys, Constit1, Constit2 ),
        in( $/shared/utt, USet ),
        in( USet, Utt ),
        $avm :: speaker( Utt, sys ),
        $avm :: consec_constits( Utt, Constit1, Constit2 ),
        $answerhood :: relevant_answer( Q, P ),
        $avm :: move( M, assert(P) )
      ],
      [ set( next_moves, set([M]) )
      ] ).
```

Listing 6.32: Selection rule for gap CR answers

Clausal Over-Answers

For polar clausal CR questions, we have already seen that the grounding rules will introduce
an action to the agenda to respond to two questions: the polar question actually asked by the
CR, and the associated MAX-QUD *wh*-question. In this case, two moves must be created, one
answering each.

```
rule( selectAnswerCR,
      [ fst( $/private/agenda, respond(Q1,Q2) ),
        $avm :: question( Q1, [], P ),
        $avm :: illoc_rel( P ),
        in( $/shared/utt, USet ),
        in( USet, Utt ),
        $avm :: speaker( Utt, sys ),
        $avm :: content( Utt, P ),
        $answerhood :: relevant_answer( Q1, P ),
        $answerhood :: relevant_answer( Q2, P ),
        $avm :: true( P, P1 ),
        $avm :: move( M1, assert(P1) ),
        $avm :: move( M2, assert(P) )
      ],
      [ set( next_moves, set([M1,M2]) )
      ] ).
```

Listing 6.33: Selection rule for polar clausal CR answers (affirmative)

As shown in listing 6.33, in the case where an answer *can* be found, this means that a
source utterance can be found, whose illocutionary content P provides an answer to both
questions directly. Two assertions can then be made, one of a new proposition $P1$ that P is
true (this will result in a primary polar answer *"Yes"*), and one asserting P itself (which will
result in the secondary, supplementary answer). If the desired system behaviour is actually
only to give bare yes/no answers in these affirmative cases, only the first assertion is added to

`next_moves` (as actually currently implemented in CLARIE).

For negative answers, the rule looks much more complex but follows the same pattern (listing 6.34): it has to first check that there is *no* source utterance which provides a positive answer to both questions, then find one that does answer the MAX-QUD *wh*-question. This utterance's content can then be asserted (giving the correct supplementary answer) together with the primary polar answer *"No"* (i.e. the proposition that the originally asked proposition *P* is not true):

```
rule( selectAnswerCR,
      [ fst( $/private/agenda, respond(Q1,Q2) ),
        $avm :: question( Q1, [], P ),
        $avm :: illoc_rel( P ),
        not (
            in( $/shared/utt, USet ) and
            in( USet, Utt ) and
            $avm :: speaker( Utt, sys ) and
            $avm :: content( Utt, P ) and
            $answerhood :: relevant_answer( Q1, P ),
            $answerhood :: relevant_answer( Q2, P )
            ),
        in( $/shared/utt, USet ),
        in( USet, Utt ),
        $avm :: speaker( Utt, sys ),
        $avm :: content( Utt, P ),
        $answerhood :: relevant_answer( Q2, P ),
        $avm :: untrue( P, P1 ),
        $avm :: move( M1, assert(P1) ),
        $avm :: move( M2, assert(P) )
      ],
      [ set( next_moves, set([M1,M2]) )
      ] ).
```

Listing 6.34: Selection rule for polar clausal CR answers (negative)

6.4.3   Asking System CRs

It is here that the choice of CR form (and reading) must be made. Firstly, there is a general choice to be made between reprise and non-reprise questions. Chapter 3 showed that non-reprise CRs were most likely to be responded to; on the other hand, the higher frequency of some reprise forms (particularly fragments and sluices), especially when clarifying NPs (which we expect to be the most frequent source of clarification) suggests that more natural behaviour might be obtained by allowing reprises. It seems sensible to allow response type to be controlled, and therefore the CR selection rules are defined as sensitive to a `reprise` flag (which can be set before starting the system). As generation (and interpretation) of reprise forms is only possible in the presence of a suitable coerced context (the grammar only licenses reprises with suitable values of MAX-QUD and SAL-UTT), the rules can differ only in whether

they coerce the context in a suitable way. Even non-reprise CRs require SAL-UTT to be coerced suitably, so this is always done – the `reprise` flag just determines whether a suitable MAX-QUD value is passed on to the generation module.

A general template for CR move selection, then is as shown in listing 6.35. Given a `clarify` action on the agenda, a new MAX-QUD and SAL-UTT are created via a contextual coercion operation, and the desired clarification question Q formed from these values and other properties of the source utterance. The `clarify` action is then removed from the agenda, the next move is set to be a move asking the new question Q, and the contextual SAL-UTT and (optionally) MAX-QUD are set, to be passed on to the generation module.

```
rule( selectClarify,
      [ fst( $/private/agenda, clarify( Utt, Constit ) ),
        $grounding :: coercion_operation( Utt, Constit, MQ, SU ),
        $avm :: clarification_question( Q, MQ, SU, Utt, Constit ),
        $avm :: move( M, ask(Q) )
      ],
      [ pop( /private/agenda ),
        set( sal_utt, SU ),
        if_do( flag( reprise, yes ), [ set( max_qud, MQ ) ] ),
        set( next_moves, set([M]) )
      ] ).
```

Listing 6.35: CR selection rule template

Note that removing the `clarify` action from the agenda does not prevent this CR from being re-asked in the future: if it is not answered, a new `clarify` action will be raised next time the system tries to ground the still pending source utterance. Note also that the new MAX-QUD and SAL-UTT values are being assigned to interface variables (visible to the generation module) but not assigned to the corresponding shared fields of the IS: this is only considered to happen once the question has actually been asked (i.e. after generation and output), and so is performed by next round of grounding/integration rules.

In the discussion of individual rules below, the use of reprise versions is assumed (as this has been one of the main purposes of the implementation). Non-reprise versions should also be possible in all cases, however: a constituent reprise fragment *"Bo?"* and a non-reprise *"Who do you mean 'Bo'?"* could both be generated from the same move.

Unknown Parameters

When clarifying unknown parameters (out-of-lexicon words or unresolvable reference), there is a choice not only of reprise or non-reprise but of the form to use – both a clausal *wh*-form (e.g. a reprise sluice) and a constituent fragment could ask suitable questions. Chapter 3 (section 3.3) suggested that sluices were common when querying definite NPs and pronouns, but that fragments were more common when querying proper names or common nouns. Verbs

were so rare as sources that we cannot be sure: for now, fragments are used.

These preferences can be specified directly by choosing which coercion operation is used based on the PoS type of the source constituent. A rule for sluices uses parameter focussing, thus creating a clausal MAX-QUD *wh*-question) and makes this the question to be asked by the next move (listing 6.36). The rule for fragments is identical except that the operation used is parameter identification (listing 6.37). The two rules are then constrained to only apply to particular phrase types.

```
rule( selectClarify,
      [ fst( $/private/agenda, clarify( Utt, Constit, unknown(P) ) ),
        $avm :: pronoun( Constit ) or $avm :: definite( Constit ),
        $grounding :: parameter_focussing( Utt, Constit, MQ, SU ),
        $avm :: move( M, ask(MQ) )
      ],
      [ pop( /private/agenda ),
        set( sal_utt, SU ),
        if_do( flag( reprise, yes ), [ set( max_qud, MQ ) ] ),
        set( next_moves, set([M]) )
      ] ).
```

Listing 6.36: Clarification rule for unknown parameters (sluices)

```
rule( selectClarify,
      [ fst( $/private/agenda, clarify( Utt, Constit, unknown(P) ) ),
        not $avm :: pronoun( Constit ) or $avm :: definite( Constit ),
        $grounding :: parameter_identification( Utt, Constit, MQ, SU ),
        $avm :: move( M, ask(MQ) )
      ],
      [ pop( /private/agenda ),
        set( sal_utt, SU ),
        if_do( flag( reprise, yes ), [ set( max_qud, MQ ) ] ),
        set( next_moves, set([M]) )
      ] ).
```

Listing 6.37: Clarification rule for unknown parameters (fragments)

Ambiguous Parameters

For ambiguous parameters (multiple possible reference), a similar choice applies: suitable CRs might again be clausal sluices or constituent fragments. Directly parallel rules can therefore be used, again allowing sluices for definites and pronouns, and fragments for others. There is a further possibility, though: an alternative question asking which of the possible referents was intended (see (Traum, 2003)'s question *"The driver or the boy?"* in section 2.4.5). As the alternatives are available as part of the clarify action specification (see section 6.3.4

above), this is of course possible given a suitable generation module, and would be derived by a selection rule as shown in listing 6.38 – however, alternative questions are not currently implemented in the grammar, so this is not used in this system.

```
rule( selectClarify,
      [ fst( $/private/agenda, clarify( Utt, Constit, ambig(P,Alts) ) ),
        $grounding :: parameter_focussing( Utt, Constit, MQ, SU ),
        $avm :: question( MQ, _Params, P ),
        $avm :: alternative_question( Q, Alts, P ),
        $avm :: move( M, ask(Q) )
      ],
      [ ...
        set( next_moves, set([M]) )
      ] ).
```

Listing 6.38: Clarification rule for ambiguous parameters (alternatives)

Inconsistent Parameters

Inconsistent parameters (those which conflict with previously held beliefs) can be queried by clausal questions in all cases: what we want to check is whether the apparently inconsistent referent which was found during grounding is really intended, so a reprise sentence or fragment is required (e.g. meaning something like *"Is it really **Paris** that you say you want to go to?"*). Again, this requires parameter focussing; the MAX-QUD question that is produced is a *wh*-question about the referent – the question to be asked is a yes/no question with this referent instantiated to the apparently inconsistent value (listing 6.39).

```
rule( selectClarify,
    [ fst( $/private/agenda, clarify( Utt, Constit, inconsistent(Param) ) ),
      $grounding :: parameter_focussing( Utt, Constit, MQ, SU ),
      $avm :: question( MQ, [Param], P ),
      $avm :: question( Q, [], P ),
      $avm :: move( M, ask(Q) )
    ],
    [ pop( /private/agenda ),
      set( sal_utt, SU ),
      if_do( flag( reprise, yes ), [ set( max_qud, MQ ) ] ),
      set( next_moves, set( [M] ) )
    ] ).
```

Listing 6.39: Clarification rule for inconsistent parameters

Inconsistent, Irrelevant and Uninterpretable Utterances

As shown in section 3.3, querying an entire sentence or utterance appears to be best done by a full reprise sentence or a conventional CR respectively. In the case of moves which have been fully interpreted and can be grounded but only in an inconsistent way, a reprise sentence seems best, as shown in listing 6.40: it is the propositional content of the move that is causing problems. In the uninterpretable and irrelevant cases, either the surface form or the intended move (the intended content of the utterance) is to be queried, and a conventional CR is used (listing 6.41 shows the lexical version for uninterpretable cases).

```
rule( selectClarify,
      [ fst( $/private/agenda, clarify( Utt, Move, inconsistent ) ),
        $grounding :: parameter_focussing( Utt, _Constit, MQ, SU ),
        $avm :: question( MQ, _Params, P ),
        $avm :: question( Q, [], P ),
        $avm :: move( M, ask(Q) )
      ],
      [ pop( /private/agenda ),
        set( sal_utt, SU ),
        if_do( flag( reprise, yes ), [ set( max_qud, MQ ) ] ),
        set( next_moves, set([M]) )
      ] ).
```

Listing 6.40: CR selection rule for inconsistent moves

```
rule( selectClarify,
      [ fst( $/private/agenda, clarify( USet ) ),
        $grounding :: lexical_identification( USet, MQ, SU ),
        $avm :: move( M, ask(MQ) )
      ],
      [ pop( /private/agenda ),
        set( sal_utt, SU ),
        set( next_moves, set([M]) )
      ] ).
```

Listing 6.41: CR selection rule for utterances

6.4.4   Grammar-based Generation

The set of moves to be generated is now passed to the generation module, which uses the grammar to produce a suitable output string which will convey that move.

Generation in HPSG can be performed using a number of methods, but its head-driven nature means it is usually particularly well suited to head-driven generation (Shieber et al., 1990). However, this does not lend itself easily to the contextually abstracted representation we now have: much of the information required to select words from the lexicon (such as

names of individuals) is now no longer in the semantic CONTENT feature, but in the con-
textually abstracted C-PARAMS set. Just as the semantic content assigned by the grammar
during parsing to an input *"John likes Mary"* will be an abstract consisting of the move
$assert(P(x, y))$ and its contextual set $\{[P : name(P, like)], [x : name(x, john)],$
$[y : name(y, mary)]\}$, when generating the grammar must take two inputs (the move and the
contextual background set of parameters) and produce an output string. This set of parame-
ters is determined by the generation module itself, given the IS background and the referents
which play roles in the desired move.

This means that generation is more straightforwardly performed by a variation of bag gen-
eration or shake-and-bake (Brew, 1992). The move and the IS background are first examined
for basic semantic units which have the required referents, and lexical items are chosen from
the lexicon on this basis. These items are used to initiate a generation chart which is extended
using a variation of the chart parser used in interpretation. Rather than spanning parts of the
input string as in parsing, the edges in this chart span a part of the input semantics, but they
are extended using exactly the same grammar. Once fully extended, any inactive edges which
span the entire input semantics are taken as possible full sign representations for the generated
sentence, and the value of the PHON attribute can then be taken as the output string.

Ellipsis in Generation

As the same grammar is being used, elliptical sentences can be generated just as they can be
interpreted, provided that a suitable context exists to license them. This has the advantage of
speeding up generation by reducing the work that must be done by the chart generator, while
also allowing more natural responses (short answers etc.), and ensuring that the interpretation
and generation capabilities are equal.

Once the intended dialogue move & background set and the possible values for MAX-QUD
and SAL-UTT have been calculated by the DME, these are passed together to the generator.
The values of MAX-QUD and SAL-UTT are then used to eliminate unnecessary entries from
the initial chart (removing the necessity to add words to cover parts of the semantics which
are already covered by these contextual variables). The chart generation process can then
produce an elliptical version of the required sentence.

This behaviour can be easily changed to prevent elliptical versions being generated: the
values of MAX-QUD and SAL-UTT are simply not passed to the generator. This also applies
to reprise CRs, of course, which require MAX-QUD values to specify their content – if the
MAX-QUD value is not passed, reprise versions cannot be generated.

6.5 Dialogue Management

Now that the input side (interpretation, grounding and IS update) and output side (selection and generation) have been described, we can start to put these together and see how the system as a whole works and deals with particular dialogue phenomena. This section starts by stepping through a simple non-clarificational dialogue in section 6.5.1, then moves onto dialogues that include user CRs in section 6.5.2 and system CRs in section 6.5.3.

6.5.1 A Basic Dialogue

To illustrate the basic principles of the system, we first consider the very simple non-clarificational dialogue shown in listing 6.42.

```
$S> Hi!
$U> hello
$S> What do you want?
$U> a ticket
$S> How do you want to travel?
$U> by plane to paris
$S> Where do you want to go from?
$U> london in march
$S> The ticket will cost 50 pounds.
$U> ok
$S> Goodbye!
```

Listing 6.42: Basic non-clarificational dialogue

Initial State

At the beginning of any interaction, the system's IS is in its default state: the SHARED part is essentially empty, and the PRIVATE part contains the default plan and associated background information. This plan contains two actions: firstly, to greet the user, and secondly, to find out what the task is (i.e. in our simple domain, to find out the answer to a particular question, the question of what the user wants $?X.want(usr, X)$). Given the abstracted representation being used, this must be specified as a question concerning a particular predicate p_1 and an associated background parameter describing that predicate: i.e. the question $?X.p_1(usr, X)$

and the parameter $p_1 : name(p_1, want)$, as shown below:[11]

$$
(324) \quad
\begin{bmatrix}
\text{PRIVATE} & \begin{bmatrix}
\text{AGENDA} & \left\langle \texttt{greet} \right\rangle \\
\text{PLAN} & \left\langle \texttt{findout}\,(?X.p_1(usr, X)) \right\rangle \\
\text{BEL} & \{\} \\
\text{BG} & \left\{ \left[p_1 : name(p_1, want) \right] \right\}
\end{bmatrix} \\
\text{SHARED} & \begin{bmatrix}
\text{COM} & \{\} \\
\text{BG} & \{\} \\
\text{QUD} & \langle\rangle \\
\text{SAL-UTT} & \langle\rangle \\
\text{UTT} & \left\langle \{\} \right\rangle \\
\text{PENDING} & \langle\rangle
\end{bmatrix}
\end{bmatrix}
$$

System Greeting

The DME now calls the `select` module, which applies the selection rules to determine the next move that the system will make. The only rule which applies to the current IS (specifically, the current state of AGENDA), is the general `selectOther` rule given in listing 6.27 above and repeated (fully instantiated) here as listing 6.43. This sets the IS variable `next_moves` to be a singleton set containing a greeting move, $\{greet\}$.

```
rule( selectOther,
      [ fst( $/private/agenda, greet ),
        $avm :: move( M, greet )
      ],
      [ set( next_moves, set([greet]) )
      ]
    ).
```

Listing 6.43: Selection rule for greetings (instantiated)

The DME now calls the `generate` module, which uses the grammar to produce a sign whose CONTENT value is a *greet* move, as in AVM (325). This sign is then assigned to the IS variable `latest_utt`, and its PHON value (its orthographic string representation *"Hi!"*) is assigned to the IS variable `inoutput`. The `output` module is then called to output the

---

[11]The IS representations here use $\{\ldots\}$ to denote a set, $\langle\ldots\rangle$ to denote a stack or stackset, and $\langle\{\ldots\}\rangle$ to denote the *nstackset* type used by the UTT record.

string to the user.

$$
(325) \quad
\begin{bmatrix}
\text{PHON} & \left\langle \text{hi} \right\rangle \\[2ex]
\text{CONT} &
\begin{bmatrix}
greet \\
\text{SPKR} & sys \\
\text{ADDR} & usr
\end{bmatrix} \\[4ex]
\text{C-PARAMS} & \left\{ \Big[sys\!:\!spkr(sys) \Big], \Big[usr\!:\!addr(usr) \Big] \right\}
\end{bmatrix}
$$

Having output the string, the DME must now update the IS accordingly: the `greet` action must be removed from the agenda now that the corresponding move has been made, and the utterance record must be updated. This is achieved by calling the `update` module to apply the update rules. Several rules apply in the current state, in sequence. First to apply is the general `initialise` rule (see listing 6.4 above), which grounds the contextual parameters corresponding to speaker and addressee (which in this case are the only two parameters, and are already actually instantiated – parameters are always instantiated for system utterances, as they are generated from known referents) and adds the utterance to PENDING and UTT, resulting in the following IS:

$$
(326) \quad
\begin{bmatrix}
\text{PRV} &
\begin{bmatrix}
\text{AGENDA} & \left\langle \texttt{greet} \right\rangle \\
\text{PLAN} & \left\langle \texttt{findout}\,(?X.p_1(usr, X)) \right\rangle \\
\text{BEL} & \{\} \\
\text{BG} & \left\{ \Big[p_1\!:\!name(p_1, want) \Big] \right\}
\end{bmatrix} \\[6ex]
\text{SHR} &
\begin{bmatrix}
\text{COM} & \{\} \\
\text{BG} & \{\} \\
\text{QUD} & \langle \rangle \\
\text{SAL-UTT} & \langle \rangle \\
\text{UTT} & \left\langle \left\{ \boxed{S_1}
\begin{bmatrix}
\text{PHON} & \left\langle \text{hi} \right\rangle \\
\text{CONT} & greet(sys, usr) \\
\text{C-PARAMS} & \left\{ sys, usr \right\}
\end{bmatrix} \right\} \right\rangle \\
\text{PENDING} & \left\langle \boxed{S_1} \right\rangle
\end{bmatrix}
\end{bmatrix}
$$

Secondly, the `integrateSysGreet` rule shown in listing 6.44 applies, which grounds the utterance as normal (removing it from the PENDING stack) and removes the top action from the agenda. The standard grounding conditions are trivially satisfied here as no uninstantiated contextual parameters will exist, and could be removed from the rule specification – for now they are left in merely to minimise the differences between rules and make the code easier to understand.

```
rule( integrateSysGreet,
     [ fst( $/shared/pending, USet ),
       in( USet, Utt ),
       $avm :: speaker( Utt, sys ),
       $avm :: get_move( Utt, lambda(Params,Move) ),
       fst( $/shared/qud, MQ ),
       fst( $/shared/sal_utt, SU ),
       $grounding :: grounding( Params, $/private/bg, $/shared/bg,
                                MQ, SU, NewShared, [] ),
       $avm :: move( Move, greet, [] )
     ],
     [ pop( /shared/pending ),
       extend( /shared/bg, NewShared ),
       pop( /private/agenda )
     ] ).
```

Listing 6.44: Grounding rule for system greetings

The resulting IS then takes the following form:

$$
(327) \quad
\begin{bmatrix}
\text{PRV} & \begin{bmatrix}
\text{AGENDA} & \langle\rangle \\
\text{PLAN} & \left\langle \texttt{findout}\,(?X.p_1(usr, X)) \right\rangle \\
\text{BEL} & \{\} \\
\text{BG} & \left\{ \begin{bmatrix} p_1 : name(p_1, want) \end{bmatrix} \right\}
\end{bmatrix} \\
\text{SHR} & \begin{bmatrix}
\text{COM} & \{\} \\
\text{BG} & \{\} \\
\text{QUD} & \langle\rangle \\
\text{SAL-UTT} & \langle\rangle \\
\text{UTT} & \left\langle \left\{ \boxed{S_1} \begin{bmatrix} \text{PHON} & \langle \text{hi} \rangle \\ \text{CONT} & greet(sys, usr) \end{bmatrix} \right\} \right\rangle \\
\text{PENDING} & \langle\rangle
\end{bmatrix}
\end{bmatrix}
$$

Finally, plan and agenda management rules apply: in this case, a `refillAgendaFromPlan` rule moves the next planned action from the plan to the agenda (as it is now the system's im-

mediate goal).

$$
(328)\quad
\begin{bmatrix}
\text{PRV} &
\begin{bmatrix}
\text{AGENDA} & \left\langle \texttt{findout}\,(?X.p_1(usr,X)) \right\rangle \\
\text{PLAN} & \langle\rangle \\
\text{BEL} & \{\} \\
\text{BG} & \left\{ \begin{bmatrix} p_1 : name(p_1, want) \end{bmatrix} \right\}
\end{bmatrix} \\[4em]
\text{SHR} &
\begin{bmatrix}
\text{COM} & \{\} \\
\text{BG} & \{\} \\
\text{QUD} & \langle\rangle \\
\text{SAL-UTT} & \langle\rangle \\
\text{UTT} & \left\langle \left\{ \boxed{S_1}\begin{bmatrix} \text{PHON} & \left\langle \text{hi} \right\rangle \\ \text{CONT} & greet(sys, usr) \end{bmatrix} \right\} \right\rangle \\
\text{PENDING} & \langle\rangle
\end{bmatrix}
\end{bmatrix}
$$

User Greeting

As with GoDiS, the system then waits for a user input. When one is received (via the `input` module), the string is passed via the `inoutput` IS variable to the `interpret` module, which parses the string using the HPSG grammar to produce a set of possible signs (assigned to the `latest_utt` variable). In this case there will be two possible parses, as shown in (329): a greeting and an utterance-anaphoric version (in other words, one which could be interpreted as a user CR querying an original utterance of *hello*):

$$
(329)\quad
\begin{bmatrix}
\text{PHON} & \left\langle \text{hello} \right\rangle \\
\text{CONT} & \begin{bmatrix} greet \\ \text{SPKR} & S \\ \text{ADDR} & A \end{bmatrix} \\
\text{C-PARAMS} & \left\{ \begin{bmatrix} S : spkr(S) \end{bmatrix}, \begin{bmatrix} A : addr(A) \end{bmatrix} \right\}
\end{bmatrix}
\quad
\begin{bmatrix}
\text{PHON} & \left\langle \text{hello} \right\rangle \\
\text{CONT} & U \\
\text{CTXT}\,|\,\text{SAL-UTT} & U\begin{bmatrix} \text{PHON} & \left\langle \text{hello} \right\rangle \end{bmatrix} \\
\text{C-PARAMS} & \left\{ \begin{bmatrix} U : sal\_utt(U) \end{bmatrix}, \\ \begin{bmatrix} S : spkr(S) \end{bmatrix}, \begin{bmatrix} A : addr(A) \end{bmatrix} \right\}
\end{bmatrix}
$$

The `update` module is now called again to ground the input and integrate it into the IS. As before, the first rule to be applied is `initialize`, grounding the speaker and addressee parameters (the *conversational participant* or *CP parameters*) and adding the utterance to PENDING and UTT. This time the grounding of CP parameters is less trivial: the parameters are not already instantiated, and the index values S, A shown above must be instantiated to their correct values usr, sys based on the knowledge that the latest speaker was the user.

The set containing both possible signs, corresponding to both possible parses, is therefore added to PENDING and UTT. Only the AVM corresponding to the parse that will end up being successfully grounded (the greeting parse) is shown here in (330) to make it easier to read –

this will be the case throughout this section:

$$
(330)\quad
\begin{bmatrix}
\text{PRV} & \begin{bmatrix}
\text{AGENDA} & \left\langle \texttt{findout}\,(?X.p_1(usr,X)) \right\rangle \\
\text{PLAN} & \langle\rangle \\
\text{BEL} & \{\} \\
\text{BG} & \left\{ \begin{bmatrix} p_1:name(p_1,want) \end{bmatrix} \right\}
\end{bmatrix} \\[4ex]
\text{SHR} & \begin{bmatrix}
\text{COM} & \{\} \\
\text{BG} & \{\} \\
\text{QUD} & \langle\rangle \\
\text{SAL-UTT} & \langle\rangle \\
\text{UTT} & \left\langle \left\{ \boxed{U_1}\begin{bmatrix} \text{PHON} & \langle hello \rangle \\ \text{CONT} & greet(usr,sys) \end{bmatrix}, \boxed{S_1}\begin{bmatrix} \text{PHON} & \langle hi \rangle \\ \text{CONT} & greet(sys,usr) \end{bmatrix} \right\} \right\rangle \\
\text{PENDING} & \left\langle \boxed{U_1} \right\rangle
\end{bmatrix}
\end{bmatrix}
$$

Now the grounding rules are tested in sequence; in this case the first rule that can apply is
`integrateUsrGreet` (as shown in listing 6.8 above). This can be satisfied by the greet-
ing version of the sign, but not by the utterance-anaphoric version as its SAL-UTT parameter
cannot be grounded in the current IS (it would require a coercion operation, as do all CRs).
For the greeting version, there are no parameters left to ground (see AVM (329) above), so
the conditions are trivially satisfied, and the effects of the greeting rule are applied, including
removing the sign from PENDING (331). As there is a previous system greeting in the utter-
ance record, only the first version of the rule (listing 6.8) can apply, so there are no further
effects; if no previous system greeting existed, the second version (listing 6.9) would apply
and a new agenda action to return the greeting would be added.

$$
(331)\quad
\begin{bmatrix}
\text{PRV} & \begin{bmatrix}
\text{AGENDA} & \left\langle \texttt{findout}\,(?X.p_1(usr,X)) \right\rangle \\
\text{PLAN} & \langle\rangle \\
\text{BEL} & \{\} \\
\text{BG} & \left\{ \begin{bmatrix} p_1:name(p_1,want) \end{bmatrix} \right\}
\end{bmatrix} \\[4ex]
\text{SHR} & \begin{bmatrix}
\text{COM} & \{\} \\
\text{BG} & \{\} \\
\text{QUD} & \langle\rangle \\
\text{SAL-UTT} & \langle\rangle \\
\text{UTT} & \left\langle \left\{ \boxed{U_1}\begin{bmatrix} \text{PHON} & \langle hello \rangle \\ \text{CONT} & greet(usr,sys) \end{bmatrix}, \boxed{S_1}\begin{bmatrix} \text{PHON} & \langle hi \rangle \\ \text{CONT} & greet(sys,usr) \end{bmatrix} \right\} \right\rangle \\
\text{PENDING} & \langle\rangle
\end{bmatrix}
\end{bmatrix}
$$

System Question

Having fully processed the user input, the DME again moves on to the `select` module to generate the next system move that will be made. This time the `selectAsk` rule applies (see listing 6.25) and an ask move relative to the agenda question is produced. As before, the `generate` and `output` modules produce a corresponding sign (AVM (332)) and output its associated string to the user.

$$(332) \quad \begin{bmatrix} \text{PHON} & \left\langle \text{what, do, you, want} \right\rangle \\[2ex] \text{CONT} & \begin{bmatrix} ask \\ \text{SPKR} & sys \\ \text{ADDR} & usr \\ \text{MSG-ARG} & \left[?X.p_1(usr, X) \right] \end{bmatrix} \\[4ex] \text{C-PARAMS} & \left\{ \left[sys : spkr(sys) \right], \left[usr : addr(usr) \right], \left[p_1 : name(p_1, want) \right] \right\} \end{bmatrix}$$

Note that as this is a system utterance containing the known predicate p_1 (which was present in the IS background), the parameter associated with this predicate already has its referent instantiated to this known value. As with all system utterances, grounding is therefore a trivial process (the system knows what the system meant). The update rules for system turns therefore always have their grounding preconditions satisfied, and reduce to GoDiS-style IS integration rules. In the case of system questions, the rule `integrateSysAsk` (see listing 6.7 above) will therefore always succeed, removing the utterance from PENDING, adding any parameters newly introduced to the common ground to SHARED/BG, and adding the newly asked question to QUD (along with adding the *wh*-phrase *what* to SAL-UTT for later ellipsis resolution). The agenda action is not removed: `findout` actions are defined only to

be removed when a question is answered.

$$
(333) \quad
\begin{bmatrix}
\text{PRV} &
\begin{bmatrix}
\text{AGENDA} & \left\langle \text{findout}\,(\boxed{Q_1}?X.p_1(usr, X))\, \right\rangle \\
\text{PLAN} & \langle\,\rangle \\
\text{BEL} & \{\} \\
\text{BG} & \left\{ \boxed{B_1}\Big[\,p_1\!:\!name(p_1, want)\,\Big] \right\}
\end{bmatrix} \\[6em]
\text{SHR} &
\begin{bmatrix}
\text{COM} & \{\} \\
\text{BG} & \left\{ \boxed{B_1} \right\} \\
\text{QUD} & \left\langle \boxed{Q_1} \right\rangle \\
\text{SAL-UTT} & \left\langle \boxed{W} \right\rangle \\
\text{UTT} & \left\langle \left\{
\boxed{S_2}\begin{bmatrix}
\text{PHON} & \langle \text{what, do, you, want} \rangle \\
\text{CONT} & ask(sys, usr, \boxed{Q_1}) \\
\text{CONSTITS} & \left\{ \ldots \boxed{W}\begin{bmatrix} \text{PHON} & \langle \text{what} \rangle \\ \text{CONTENT} & X \end{bmatrix} \ldots \right\}
\end{bmatrix}
\right\}, \left\langle \boxed{U_1}\Big[\text{PHON}\ \langle \text{hello} \rangle\Big], \boxed{S_1}\Big[\text{PHON}\ \langle \text{hi} \rangle\Big] \right\rangle \right\rangle \\
\text{PENDING} & \langle\,\rangle
\end{bmatrix}
\end{bmatrix}
$$
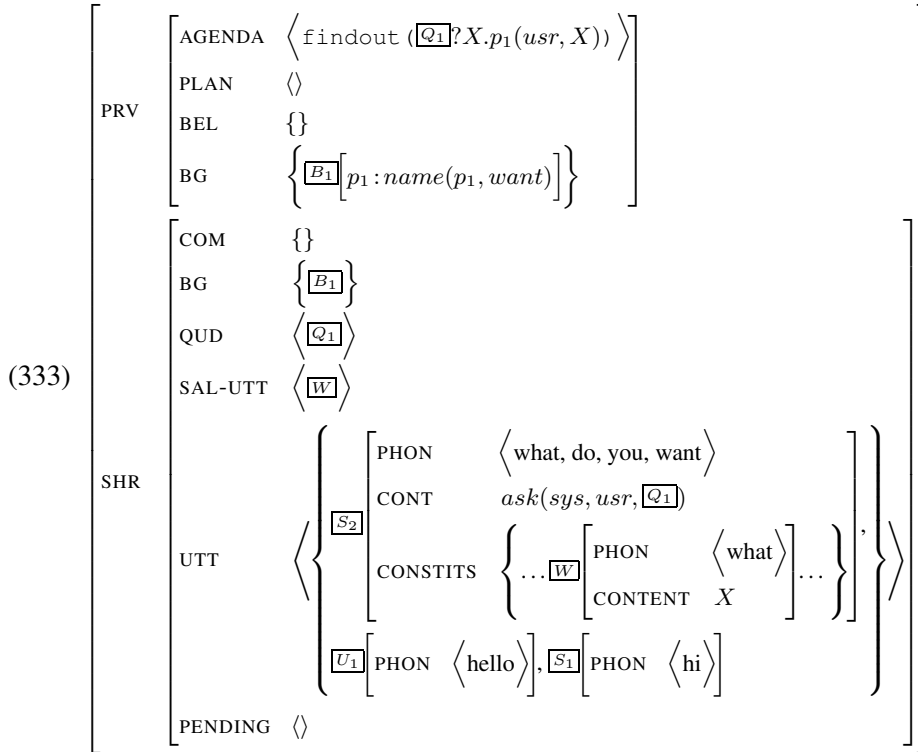
User Answer

For the next user input, the elliptical answer *"a ticket"*, the grammar produces a highly con-
textualised sign as shown in AVM (334): a declarative fragment whose propositional content
derives from context (from the value of MAX-QUD) but which is quantified over by a variable
x_1, where x_1 is defined by the logical relation R_1 (the indefinite quantifier) and the predicate
P_2 (the property of being a ticket). The exact role of x_1 in the proposition is again determined
by context, by associating it with the referent of the value of SAL-UTT. All of these contextual

variables must be identified in context by the grounding process.

$$
(334) \quad
\begin{bmatrix}
\text{PHON} & \left\langle a, ticket \right\rangle \\[1em]
\text{CONT} & \begin{bmatrix}
ask \\
\text{SPKR} & S \\
\text{ADDR} & A \\
\text{MSG-ARG} & \begin{bmatrix}
\text{QUANTS} & \left\{ \left[x_1 : R_1(x_1, P_2) \right] \right\} \\
\text{NUCL} & \boxed{1}
\end{bmatrix}
\end{bmatrix} \\[3em]
\text{CTXT} & \begin{bmatrix}
\text{MAX-QUD} & \boxed{3}\left[\text{PROP} \,|\, \text{NUCL} \quad \boxed{1} \right] \\
\text{SAL-UTT} & \boxed{4}\left[\text{INDEX} \quad x_1 \right]
\end{bmatrix} \\[2em]
\text{C-PARAMS} & \left\{
\begin{array}{l}
\left[S : spkr(S) \right], \left[U : addr(U) \right], \\
\left[R_1 : R_1 = exist \right], \left[P_2 : name(P_2, ticket) \right], \\
\left[\boxed{3}QUD : max\_qud(QUD) \right], \left[\boxed{4}UTT : sal\_utt(UTT) \right]
\end{array}
\right\}
\end{bmatrix}
$$

The first two parameters (the CP parameters) are identified by the `initialize` rule as before. The remaining parameters must be assigned by the `integrateAnswer` rule (see listing 6.5), via the `grounding` condition: the relation R_1 and predicate P_2 are unambiguously found by lexicon lookup, and the MAX-QUD and SAL-UTT values taken from the top of the QUD and SAL-UTT stacks. As the current maximal QUD is a suitable *wh*-question, the grounding succeeds with these values to produce a fully specified sign, with the content being the proposition that the user wants a ticket. This is trivially consistent with the current shared commitments (as there are none), so the consistency check also succeeds and the rule can apply.

The effects of the rule are to remove the grounded utterance from PENDING, remove the answered question from QUD (with its associated *wh*-phrase from SAL-UTT) and to add the proposition to COM together with a second proposition that it resolves the question. The background is also extended to include newly introduced parameters, including (using our

simple protocol for existentially quantified parameters) those in QUANTS.

$$
(335) \begin{bmatrix}
\text{PRV} & \begin{bmatrix}
\text{AGENDA} & \left\langle \texttt{findout}\,(\boxed{Q_1}?X.p_1(usr, X)) \right\rangle \\
\text{PLAN} & \langle\rangle \\
\text{BEL} & \{\} \\
\text{BG} & \left\{ \boxed{B_1}\Big[p_1 : name(p_1, want)\Big] \right\}
\end{bmatrix} \\[2em]
\text{SHR} & \begin{bmatrix}
\text{COM} & \left\{ \boxed{P_1}, resolves(\boxed{P_1}, \boxed{Q_1}) \right\} \\
\text{BG} & \left\{ \boxed{B_1}, \Big[x_1 : r_1(x_1, p_2)\Big], \Big[r_1 : r_1 = exist\Big], \Big[p_2 : name(p_2, ticket)\Big] \right\} \\
\text{QUD} & \langle\rangle \\
\text{SAL-UTT} & \langle\rangle \\
\text{UTT} & \left\langle \left\{ \begin{matrix}
\boxed{U_2}\begin{bmatrix} \text{PHON} & \langle a, ticket \rangle \\ \text{CONT} & \boxed{P_1}\exists x_1.p_1(usr, x_1) \end{bmatrix}, \boxed{S_2}\Big[\text{PHON} & \langle what, do, you, want \rangle\Big], \\
\boxed{U_1}\Big[\text{PHON} & \langle hello \rangle\Big], \boxed{S_1}\Big[\text{PHON} & \langle hi \rangle\Big]
\end{matrix} \right\} \right\rangle \\
\text{PENDING} & \langle\rangle
\end{bmatrix}
\end{bmatrix}
$$

Plan Management

Some plan management update rules now also come into effect, removing the `findout` action from the agenda now that COM contains the proposition that the question is resolved, and then loading a new plan on the basis of the fact that the user wants a ticket (the only plans in the simple domain are to book a ticket (or flight/trip/etc.) or to book a room).

This plan contains a number of new actions, firstly to find out the answer to certain questions such as place of departure, destination and method of travel, then to look up a price in a (trivial) database) and inform the user, and finally to close the conversation. The first of these actions is placed on the (now empty) agenda. This leads to a new system move being selected and a corresponding question *"How do you want to travel?"* being generated and output, and the rest of the dialogue proceeds as outlined above. In general, this management schema follows GoDiS.

However, a few points are worth mentioning. Firstly, as the utterance record is of finite length 4, the next system utterance will, when integrated, cause the first system greeting to drop off the end and no longer be available (e.g. as an antecedent for clarifications). Secondly, the *accommodation* rules which are one of the key features of GoDiS are kept, allowing questions on the plan to be answered without their being explicitly asked: in our example dialogue, the user input *"by plane to paris"* is treated as two fragments, one of which is an elliptical answer to the explicit question *"How do you want to travel?"*, and one of which is an elliptical answer to the unasked, but planned, question *"Where do you want to go?"*.

Accommodation is achieved by a set of rules which take a question from the agenda or plan, check that the current utterance can be grounded as a relevant answer to it, and remove it from agenda or plan and add instead to QUD as if it had been explicitly asked. The standard grounding/integration rules can then apply as with standard questions.

```
rule( accommodatePlanQuestion,
     [ fst( $/shared/pending, USet ),
       in( USet, Utt ),
       $avm :: speaker( Utt, usr ),
       $avm :: get_move( Utt, lambda(Params,Move) ),
       in( $/private/plan, raise(Q) ),
       $avm :: wh_phrase( Q, SU ),
       $grounding :: grounding( Params, $/private/bg, $/shared/bg,
                                Q, SU, _NewShared, [] ),
       $grounding :: consistent( Move ),
       $grounding :: consistent( Move, $/shared/com ),
       $avm :: move( Move, assert(P) ),
       $answerhood :: relevant_answer( Q, P )
     ],
     [ del( /private/plan, raise(Q) ),
       push( /shared/qud, Q ),
       push( /shared/sal_utt, SU )
     ] ).
```

Listing 6.45: Accommodation rule for plan questions

6.5.2   User Clarifications

Moving from this kind of simple dialogue to one that involves CRs asked by the user now requires no more than the use of the CR grounding rules already described in section 6.3.2, together with a method of generating suitable answers. Some examples of possible user CR dialogue excerpts are shown in listing 6.46 as illustration; firstly a conventional lexical CR, secondly a clausal reprise fragment, and finally a lexical reprise gap.

```
Lexical (conventional):
...
$S> What do you want?
$U> pardon
$S> What do you want?
...

Clausal (fragment, correct):
...
$S> The ticket will cost 50 pounds.
$U> fifty?
$S> Yes.
...

Clausal (fragment, incorrect):
...
$S> The ticket will cost 50 pounds.
$U> sixty?
$S> No, 50.
...

Constituent (fragment, noun or indefinite):
...
$S> Do you want to book a trip?
$U> trip? / a trip?
$S> Flight.
...

Lexical (gap):
...
$S> Where do you want to go from?
$U> to ...
$S> Go from.
...
```

Listing 6.46: Example dialogue excerpts with user clarification

The basic process follows the same schema for all these examples. The first system utterance is processed as normal, giving an IS in which it has been grounded (by the system at least):

$$(336) \quad \begin{bmatrix} \text{AGENDA} & \langle \dots \rangle \\ \text{COM} & \{ \dots \} \\ \text{QUD} & \langle \dots \rangle \\ \text{UTT} & \left\langle \left\{ \boxed{S_1}\left[\text{PHON} \quad \langle \dots \rangle \right], \dots \right\} \right\rangle \\ \text{PENDING} & \langle \rangle \end{bmatrix}$$

The user utterance is then added to pending while grounding rules are tested:

$$(337) \quad \begin{bmatrix} \text{AGENDA} & \langle \dots \rangle \\ \text{COM} & \{ \dots \} \\ \text{QUD} & \langle \dots \rangle \\ \text{UTT} & \left\langle \left\{ \boxed{U_1}\left[\text{PHON} \quad \langle \dots \rangle \right], \boxed{S_1}\left[\text{PHON} \quad \langle \dots \rangle \right], \dots \right\} \right\rangle \\ \text{PENDING} & \left\langle \boxed{U_1} \right\rangle \end{bmatrix}$$

The user utterance then becomes grounded as a clarification request by one of the update rules (possibly involving coercion of a new QUD), and its resulting content is a question concerning a constituent of the source utterance (the system's original utterance). This question is added to QUD, the user utterance removed from PENDING, and the resulting action is for the system to respond to this question:

$$(338) \quad \begin{bmatrix} \text{AGENDA} & \left\langle \texttt{respond}(\boxed{Q}), \dots \right\rangle \\ \text{COM} & \{ \dots \} \\ \text{QUD} & \left\langle \boxed{Q}, \dots \right\rangle \\ \text{UTT} & \left\langle \left\{ \boxed{U_1}\begin{bmatrix} \text{PHON} & \langle \dots \rangle \\ \text{CONT} & \boxed{Q}\left[\dots \boxed{X} \dots \right] \end{bmatrix}, \boxed{S_1}\begin{bmatrix} \text{PHON} & \langle \dots \rangle \\ \text{CONSTITS} & \{ \dots \boxed{X} \dots \} \end{bmatrix}, \dots \right\} \right\rangle \\ \text{PENDING} & \langle \rangle \end{bmatrix}$$

The selection rules then produce a suitable system response, which discharges the agenda action and removes the clarification question from QUD as it has been answered (also adding

a new shared belief to COM):

$$(339)\quad \begin{bmatrix} \text{AGENDA} & \langle \dots \rangle \\ \text{COM} & \{\dots P(\boxed{X}) \dots\} \\ \text{QUD} & \langle \dots \rangle \\ \text{UTT} & \left\langle \left\{ \boxed{S_2}\begin{bmatrix} \text{PHON} & \langle \dots \rangle \\ \text{CONT} & P(\boxed{X}) \end{bmatrix}, \boxed{U_1}\begin{bmatrix} \text{PHON} & \langle \dots \rangle \end{bmatrix}, \boxed{S_1}\begin{bmatrix} \text{PHON} & \langle \dots \rangle \end{bmatrix}, \dots \right\} \right\rangle \\ \text{PENDING} & \langle \rangle \end{bmatrix}$$

Note that the IS is acting as a model of the context from the point of view of the system. It is necessarily the case that during clarificational dialogue, the views of the context of the two participants must differ: when the system produces an utterance, it is aware of its content; if the user cannot instantiate that content, their contexts differ and clarification ensues – which can only be resolved by the system using its own view of the context, including what was said and what was intended to be meant. The SHARED part of the IS is therefore an optimistic view of the common ground from the system's point of view. Explicit modelling of the user's context might be possible but does not appear necessary: interpretation of a user utterance as a CR and its subsequent discussion does not require this.

Conventional CR Example

Like almost all CRs, a conventional CR such as *"Pardon?"* cannot be handled by the standard `integrateUsrAsk` rule as it requires the source utterance to be specified via a coercion operation in order to ground its SAL-UTT parameter (see section 6.3.2 above). This is performed by the rule `integrateUsrCR`. The non-reprise version of this rule can apply (listing 6.11) as conventional CRs have their content specified by the grammar and do not require a coerced MAX-QUD question to reconstruct elliptical meaning.

Before application of the rule, the relevant parts of the IS are as shown in AVM (340)

below. The new CR utterance is in PENDING, ready for grounding.

$$(340) \quad \begin{bmatrix} \text{AGENDA} & \langle \dots \rangle \\ \text{COM} & \{ \dots \} \\ \text{QUD} & \langle \dots \rangle \\ \text{UTT} & \left\langle \left\{ \boxed{U_1} \begin{bmatrix} \text{PHON} & \langle \text{pardon} \rangle \\ \text{CONT} & \left[ask(usr, ?X.utter(usr, X)) \right] \\ \text{C-PARAMS} & \left\{ \left[X : sal\_utt(X) \right] \right\} \end{bmatrix}, \\ \boxed{S_1} \left[\text{PHON} \quad \langle \text{what, do, you, want} \rangle \right], \dots \right\} \right\rangle \\ \text{PENDING} & \left\langle \boxed{U_1} \right\rangle \end{bmatrix}$$

Grounding requires only the CPs to be (trivially) instantiated via `initialize`, and the latest system utterance to be identified and unified with the SAL-UTT parameter by `integrateUsrCR`; as the only restriction on this feature is that its value be a complete utterance (a sign of type *root-cl* – see AVM (297) above), this will always succeed, causing the user CR utterance to be removed from PENDING, the newly asked clarification question added to QUD, and an action to respond to the question added to the agenda.

$$(341) \quad \begin{bmatrix} \text{AGENDA} & \left\langle \texttt{respond}(\boxed{Q_1}), \dots \right\rangle \\ \text{COM} & \{ \dots \} \\ \text{QUD} & \left\langle \boxed{Q_1}, \dots \right\rangle \\ \text{UTT} & \left\langle \left\{ \boxed{U_1} \begin{bmatrix} \text{PHON} & \langle \text{pardon} \rangle \\ \text{CONT} & \left[ask(usr, \boxed{Q_1}[?\boxed{S_1}.utter(sys, \boxed{S_1})]) \right] \end{bmatrix}, \\ \boxed{S_1} \left[\text{PHON} \quad \langle \text{what, do, you, want} \rangle \right], \dots \right\} \right\rangle \\ \text{PENDING} & \langle \rangle \end{bmatrix}$$

The system response is produced by the selection rule `selectAnswerCR` (see listing 6.31 above), and in this particular case the next move selected is to assert the proposition that the system uttered the previously recorded utterance. The generation module then uses the grammar to generate a sign which has this move as its content (in this case, an elliptical utterance-anaphoric bare answer referring to the previous utterance and having the same surface form). This is output and grounded in the standard manner (integrated as an answer using the rule `integrateSysAnswer`, removing the clarification question from QUD and the `respond` action from the agenda, and adding appropriate propositions to the

shared commitments).

$$
(342) \quad
\begin{bmatrix}
\text{AGENDA} & \langle \dots \rangle \\[4pt]
\text{COM} & \left\{ \boxed{P_1}, \, resolves(\boxed{P_1}, \boxed{Q_1}), \dots \right\} \\[4pt]
\text{QUD} & \langle \dots \rangle \\[6pt]
\text{UTT} &
\left\langle
\left\{
\begin{array}{l}
\boxed{S_2}
\begin{bmatrix}
\text{PHON} & \langle \text{what, do, you, want} \rangle \\
\text{CONT} & \begin{bmatrix} assert(sys, \boxed{P_1} utter(sys, \boxed{S_1})) \end{bmatrix}
\end{bmatrix}, \\[18pt]
\boxed{U_1}
\begin{bmatrix}
\text{PHON} & \langle \text{pardon} \rangle \\
\text{CONT} & \begin{bmatrix} ask(usr, \boxed{Q_1}?\boxed{S_1}.utter(sys, \boxed{S_1})) \end{bmatrix}
\end{bmatrix}, \\[18pt]
\boxed{S_1}\begin{bmatrix} \text{PHON} & \langle \text{what, do, you, want} \rangle \end{bmatrix}, \dots
\end{array}
\right\}
\right\rangle \\[6pt]
\text{PENDING} & \langle \rangle
\end{bmatrix}
$$

Clausal CR Examples

Correct Example The first clausal CR example *"fifty?"* follows the same overall pattern. The main exception is that the user CR is elliptical (a reprise fragment) and therefore its interpretation requires a further contextual coercion operation to provide the necessary MAX-QUD. Before grounding, the pending utterance is also ambiguous: the interpretation module produces a set of possible signs, including the clausal and constituent CR versions shown in AVMs (343) and (344):

$$
(343) \quad
\begin{bmatrix}
\text{PHON} & \langle \text{fifty} \rangle \\[4pt]
\text{CONT} \,|\, \text{MSG-ARG} & \begin{bmatrix} ?.\boxed{1} \end{bmatrix} \\[4pt]
\text{HEAD-DTR} \,|\, \text{CONT} & \boxed{3}\begin{bmatrix} R : R = 50 \end{bmatrix} \\[6pt]
\text{CTXT} &
\begin{bmatrix}
\text{SAL-UTT} & \boxed{S}\begin{bmatrix} \text{CONT} & \boxed{2}\begin{bmatrix} \text{INDEX} & R \end{bmatrix} \end{bmatrix} \\[8pt]
\text{MAX-QUD} & \boxed{Q}\begin{bmatrix} ?\boxed{2},\boxed{1}\begin{bmatrix} illoc\text{-}rel \end{bmatrix} \end{bmatrix}
\end{bmatrix} \\[10pt]
\text{C-PARAMS} & \left\{ \boxed{3}, \boxed{S}, \boxed{Q} \right\}
\end{bmatrix}
$$

$$
(344) \quad
\begin{bmatrix}
\text{PHON} & \boxed{1}\langle \text{fifty} \rangle \\[4pt]
\text{CONT} \,|\, \text{MSG-ARG} & \boxed{Q} \\[4pt]
\text{HEAD-DTR} \,|\, \text{CONT} & \boxed{S} \\[6pt]
\text{CTXT} &
\begin{bmatrix}
\text{SAL-UTT} & \boxed{S}\begin{bmatrix} \text{PHON} & \boxed{1} \end{bmatrix} \\[8pt]
\text{MAX-QUD} & \boxed{Q}\begin{bmatrix} ?\boxed{2}.spkr\_meaning\_rel(S, \boxed{S}, \boxed{2}) \end{bmatrix}
\end{bmatrix} \\[10pt]
\text{C-PARAMS} & \left\{ \boxed{S}, \boxed{Q} \right\}
\end{bmatrix}
$$

There will also be an utterance-anaphoric lexical version as a possible interpretation, but

this will not get used in this case. As can be seen above, the constituent and clausal versions have very different specifications on the values to which MAX-QUD and SAL-UTT can be instantiated in grounding.

The various versions of the `integrateUsrCR` rule (which apply the different coercion operations) are tested in turn until one succeeds. First, the non-reprise version (as used in the previous section) fails, as all of the possible pending signs have a MAX-QUD contextual parameter which is constrained to be some kind of CR question, and which therefore cannot be grounded to the current maximal member of QUD. Second, the constituent CR version of the rule is tried (listing 6.12), which would produce a MAX-QUD of the kind required by AVM (344). However, this fails, as the source constituent would have to be a determiner (which is ruled out in the grounding preconditions, as constituent CRs appear only to query content words). The third version to be tried is the clausal CR version (listing 6.13), and this succeeds as number determiners are allowed as sources by this rule: the coerced value of MAX-QUD becomes the question *"For which R are you telling me it will cost R pounds?"*. This can be unified with the version in AVM (343), and grounding now succeeds: grounding the parameter associated with the word *fifty* instantiates R to the relation r_1 in the IS background; and the grounded content of the fragment becomes *"Is it **fifty**$_R$ that you are telling me it will cost R pounds?"*;

A further possibility for grounding might have been to take the utterance as a (lexical) reprise gap rather than a clausal fragment. However, the ordering of the grounding rules is such that the fragment interpretation is preferred - only if this had failed (e.g. if the determiner in question was not numerical and therefore very unlikely to be the subject of either constituent or clausal CR) would the gap rule have applied.

$$
(345) \quad \begin{bmatrix}
\text{AGENDA} & \left\langle \texttt{respond}\,(\boxed{Q_1},\boxed{Q_2}),\ldots \right\rangle \\[4pt]
\text{COM} & \{\ldots\} \\[4pt]
\text{BG} & \left\{ \left[p_1:name(p_1,cost)\right], \left[r_1:r_1=50\right], \left[p_2:name(p_2,pound)\right] \right\} \\[4pt]
\text{QUD} & \left\langle \boxed{Q_1},\ldots \right\rangle \\[4pt]
\text{UTT} & \left\langle \left\{ \begin{aligned} &\boxed{U_1}\begin{bmatrix} \text{PHON} & \left\langle \text{fifty} \right\rangle \\ \text{CONT} & \left[ask(usr,\boxed{Q_1}?.assert(sys,\exists\{x_1:r_1(x_1,p_2)\}.p_1(x_1))) \right] \\ \text{MAX-QUD} & \left[\boxed{Q_2}?R.assert(sys,\exists\{x_1:R(x_1,p_2)\}.p_1(x_1)) \right] \end{bmatrix}, \\[8pt] &\boxed{S_1}\begin{bmatrix} \text{PHON} & \left\langle \text{it, will, cost, 50, pounds} \right\rangle \\ \text{CONT} & \left[\boxed{M_1}assert(sys,\exists\{x_1:r_1(x_1,p_2)\}.p_1(x_1)) \right] \end{bmatrix},\ldots \end{aligned} \right\} \right\rangle \\[4pt]
\text{PENDING} & \langle\,\rangle
\end{bmatrix}
$$

The selection rules, together with the `answerhood` module, can now produce an af-

firmative answer based on the current IS (the content of the source utterance, as previously grounded by the system, is indeed the content under question – the CR question's queried propositional content is identical to that of the source utterance). Given that the current maximal QUD is a question which is being answered affirmatively, the generation module licenses an elliptical polar answer *"Yes"*.

Incorrect Example   In the second example *"sixty?"*, the grounding of the user CR proceeds in very much the same way as above (although in this case, there is a further reason why the constituent CR update rule could not apply: no antecedent utterance with the same phonology could be found to ground the SAL-UTT parameter). The IS after grounding (AVM (346)) looks very similar to AVM (345), except that the CR question now asks about a proposition which is *not* identical to the content of the source utterance (the existentially quantified parameter contains the logical relation $r_2 = 60$ instead of $r_1 = 50$).[12]

$$(346) \begin{bmatrix} \text{AGENDA} & \left\langle \texttt{respond}(\boxed{Q_1}, \boxed{Q_2}), \dots \right\rangle \\[4pt] \text{COM} & \{\dots\} \\[4pt] \text{BG} & \left\{ \begin{bmatrix} p_1 : name(p_1, cost) \end{bmatrix}, \begin{bmatrix} r_1 : r_1 = 50 \end{bmatrix}, \begin{bmatrix} p_2 : name(p_2, pound) \end{bmatrix}, \\ \begin{bmatrix} r_2 : r_2 = 60 \end{bmatrix} \right\} \\[4pt] \text{QUD} & \left\langle \boxed{Q_1}, \dots \right\rangle \\[4pt] \text{UTT} & \left\langle \left\{ \begin{matrix} \boxed{U_1} \begin{bmatrix} \text{PHON} & \left\langle \text{sixty} \right\rangle \\ \text{CONT} & \begin{bmatrix} ask(usr, \boxed{Q_1}?.assert(sys, \exists\{x_1 : r_2(x_1, p_2)\}.p_1(x_2))) \end{bmatrix} \\ \text{MAX-QUD} & \begin{bmatrix} \boxed{Q_2}?R.assert(sys, \exists\{x_1 : R(x_1, p_2)\}.p_1(x_2)) \end{bmatrix} \end{bmatrix}, \\[4pt] \boxed{S_1} \begin{bmatrix} \text{PHON} & \left\langle \text{it, will, cost, 50, pounds} \right\rangle \\ \text{CONT} & \begin{bmatrix} assert(sys, \exists\{x_1 : r_1(x_1, p_2)\}.p_1(x_1)) \end{bmatrix} \end{bmatrix}, \dots \end{matrix} \right\} \right\rangle \\[4pt] \text{PENDING} & \langle\rangle \end{bmatrix}$$

The selection rules and answerhood module cannot therefore produce an affirmative answer as before, and the negative answer version is used instead (listing 6.34 above). This means that both questions in the `respond` action must be answered, firstly with a negative answer *"No"* and secondly with the elliptical fragment *"50"*.

---

[12]Note that the content of the CR is assigned via a coerced MAX-QUD, which is in turn formed from the source utterance by abstraction of the parameter associated with the relation under question r_1. In Prolog terms, this abstraction must be achieved by replacing r_1 with an uninstantiated variable in the MAX-QUD question, which then becomes re-instantiated during grounding of the CR to the suggested r_2, thus avoiding the accidental (and undesirable) unification of r_1 with r_2.

Constituent CR Examples

Noun Example The first constituent CR example *"trip?"* is straightforward and follows the standard pattern. The utterance is again ambiguous, and clausal, constituent and lexical versions are produced by the grammar. This time, the update rule for constituent CRs can apply: phonological parallelism constraints are satisfied, and the constituent under question is a content word (a CN). The question asked by the CR is therefore resolved as being a question about the intended content of the source constituent (the parallel fragment *trip* in the original system utterance):

$$
(347)\quad
\begin{bmatrix}
\text{AGENDA} & \left\langle \texttt{respond}(\boxed{Q_1}), \dots \right\rangle \\[4pt]
\text{COM} & \{ \dots \} \\[4pt]
\text{BG} & \left\{ \left[p_1 : name(p_1, book) \right], \left[r_1 : r_1 = exist \right], \boxed{X_1}\left[p_2 : name(p_2, trip) \right] \right\} \\[4pt]
\text{QUD} & \left\langle \boxed{Q_1}, \dots \right\rangle \\[4pt]
\text{UTT} & \left\langle \left\{
\boxed{U_1}\begin{bmatrix}
\text{PHON} & \left\langle trip \right\rangle \\
\text{CONT} & \left[ask(usr, \boxed{Q_1}?X.spkr\_meaning\_rel(sys, \boxed{C_1}, X)) \right] \\
\text{MAX-QUD} & \boxed{Q_1}
\end{bmatrix},
\boxed{S_1}\begin{bmatrix}
\text{PHON} & \left\langle \dots, book, a, trip \right\rangle \\
\text{CONT} & \left[ask(sys, \dots) \right] \\
\text{CONSTITS} & \left\{ \dots, \boxed{C_1}\begin{bmatrix} \text{PHON} & \left\langle trip \right\rangle \\ \text{CONT} & \boxed{X_1} \end{bmatrix}, \dots \right\}
\end{bmatrix}, \dots
\right\} \right\rangle \\[4pt]
\text{PENDING} & \langle \rangle
\end{bmatrix}
$$

Selection rules can now produce a relevant answer: as the current QUD is suitable, this can be an elliptical fragment with the intended content (the predicate p_2 originally intended to be conveyed by the word *trip*). The obvious and extremely unhelpful answer *"trip"* is avoided as the selection module tries to choose a different description of the queried predicate if possible – in this case, a synonymous word *flight* can be found in the domain model and is used instead.

Indefinite Example The same process would apply in the case of a definite NP question, where the CR is querying the reference of that definite NP. In the case of an indefinite, however, the fragment cannot be grounded as a CR in the standard way, as there is no contextual parameter associated with the source indefinite NP as a whole (i.e. indefinites cannot be the source of clarification – see chapter 4).[13] However, it can be interpreted as a focussed CR asking a question about the noun; as this rule is preferred to the only other possibility (a gap),

---

[13]The possibility of clarification of specific/definite indefinites is mentioned in chapter 4 - this is ignored in the current system as it appears to be so rare.

it applies and this interpretation is assigned. The alternative of a CR with focus on the determiner is grammatically possible, but cannot be grounded by the update rules, as the indefinite determiner is not taken as a possible source of clarification.

Given this, the dialogue proceeds exactly as in the simple CN example above, and the resulting system response is therefore the same. Note that while this response *"Flight"* is probably acceptable, a more natural response seems to be one which maintains some syntactic parallelism with the original question, *"A flight"*. This could be achieved via a new selection rule for responding to focussed CRs, but this has not yet been implemented.

Lexical Gap Example

The final example *"to ..."* receives its interpretation as a reprise gap because all other grounding rules fail: the complementiser *to* is a function word which cannot be taken as the source of clausal or constituent questions.

$$
(348)\quad
\begin{bmatrix}
\text{AGENDA} & \left\langle \texttt{respond}(\boxed{Q_1}),\dots \right\rangle \\[4pt]
\text{COM} & \{\dots\} \\[4pt]
\text{BG} & \{\dots\} \\[4pt]
\text{QUD} & \left\langle \boxed{Q_1},\dots \right\rangle \\[8pt]
\text{UTT} & \left\langle
\begin{Bmatrix}
\boxed{U_1}
\begin{bmatrix}
\text{PHON} & \left\langle \text{to} \right\rangle \\
\text{CONT} & \left[ask(usr,\boxed{Q_1}?C.utter\_consec(sys,\boxed{C_1},C)) \right]
\end{bmatrix}, \\[12pt]
\boxed{S_1}
\begin{bmatrix}
\text{PHON} & \left\langle \dots,\text{to, go, from} \right\rangle \\
\text{CONT} & \left[ask(sys,\dots) \right] \\
\text{CONSTITS} & \left\{ \dots, \boxed{C_1}\!\left[\text{PHON}\ \left\langle \text{to} \right\rangle\right], \boxed{C_2}\!\left[\text{PHON}\ \left\langle \text{go, from} \right\rangle\right],\dots \right\}
\end{bmatrix},\dots
\end{Bmatrix}
\right\rangle \\[8pt]
\text{PENDING} & \langle\rangle
\end{bmatrix}
$$
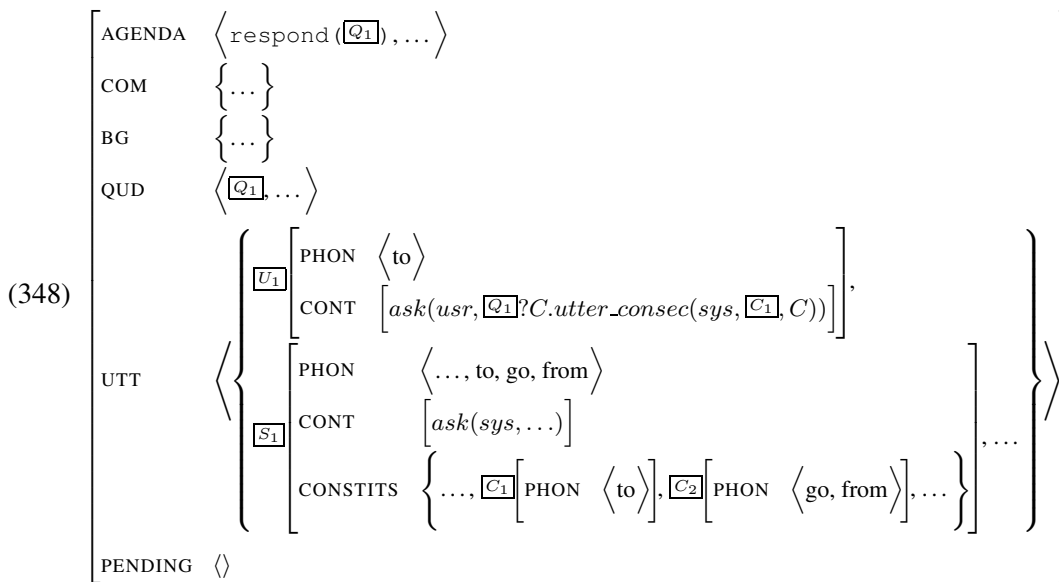
The answer to this CR question (of what was uttered after the word *to*) is obtained straightforwardly from the utterance record, and the response is generated accordingly.

6.5.3 System Clarifications

Examples of dialogue excerpts with system-generated CRs are shown in listing 6.47: a constituent fragment used to query an out-of-vocabulary word, a clausal sluice used to query ambiguous reference, and a clausal fragment used to query apparent inconsistency. All follow the same overall schema, which is outlined here. The trigger for this kind of dialogue is a user utterance which cannot be grounded due to some problematic contextual parameter. The

```
Constituent fragment (out-of-lexicon noun):
...
$S> How do you want to travel?
$U> by pullman
$S> 'Pullman'?
$U> train
$S> When do you want to leave?
...


Clausal sluice (ambiguity of referent):
...
$U> i want to go to that city
$S> Which city?
$U> paris
$S> How do you want to travel?
...


Clausal fragment (inconsistency):
...
$U> i want to go to paris
$S> How do you want to travel?
$U> no i want to go to london
$S> london?
$U> yes london
$S> How do you want to travel?
```

Listing 6.47: Example dialogue excerpts with system clarification

user utterance is added to PENDING while grounding rules are tested:

$$
(349) \quad
\begin{bmatrix}
\text{AGENDA} & \left\langle \ldots \right\rangle \\[4pt]
\text{COM} & \left\{ \ldots \right\} \\[4pt]
\text{BG} & \left\{ \ldots \right\} \\[4pt]
\text{QUD} & \left\langle \ldots \right\rangle \\[4pt]
\text{UTT} & \left\langle \left\{ \boxed{U_1}
\begin{bmatrix}
\text{PHON} & \left\langle \ldots \right\rangle \\
\text{C-PARAMS} & \left\{ \ldots, \boxed{X_1}, \ldots \right\}
\end{bmatrix}, \ldots \right\} \right\rangle \\[4pt]
\text{PENDING} & \left\langle \boxed{U_1} \right\rangle
\end{bmatrix}
$$

Due to the problematic parameter, the standard `integrate` successful grounding rules
(section 6.3.2) all fail, and instead the `clarify` unsuccessful grounding rules (section 6.3.4)
apply, leaving the utterance in the pending stack, and adding an agenda action to clarify this

parameter:

$$
(350) \quad
\begin{bmatrix}
\text{AGENDA} & \left\langle \texttt{clarify}(\boxed{X}), \ldots \right\rangle \\
\text{COM} & \{\ldots\} \\
\text{BG} & \{\ldots\} \\
\text{QUD} & \langle \ldots \rangle \\
\text{UTT} & \left\langle \left\{ \boxed{U_1}
\begin{bmatrix}
\text{PHON} & \langle \ldots \rangle \\
\text{C-PARAMS} & \{\ldots, \boxed{X}, \ldots\}
\end{bmatrix}, \ldots \right\} \right\rangle \\
\text{PENDING} & \left\langle \boxed{U_1} \right\rangle
\end{bmatrix}
$$

A suitable CR is generated and output, removing the `clarify` action from the agenda. Standard update rules now apply, adding the newly asked CR question to QUD.

$$
(351) \quad
\begin{bmatrix}
\text{AGENDA} & \langle \ldots \rangle \\
\text{COM} & \{\ldots\} \\
\text{BG} & \{\ldots\} \\
\text{QUD} & \left\langle \boxed{Q}, \ldots \right\rangle \\
\text{UTT} & \left\langle \left\{ \boxed{S_1}
\begin{bmatrix}
\text{PHON} & \langle \ldots \rangle \\
\text{CONT} & \boxed{Q}[\ldots \boxed{X} \ldots]
\end{bmatrix}, \boxed{U_1}\left[\text{PHON} \ \langle \ldots \rangle\right], \ldots \right\} \right\rangle \\
\text{PENDING} & \left\langle \boxed{U_1} \right\rangle
\end{bmatrix}
$$

The next user turn is now processed, first being added to PENDING as usual while it is grounded. As shown here, we assume that it provides an answer to the system CR. If not, standard dialogue processing rules will apply, but the problematic user utterance will remain in PENDING, and at some point the CR may be re-asked (if it is the only pending utterance and still part of the UTT utterance record).

$$
(352) \quad
\begin{bmatrix}
\text{AGENDA} & \langle \ldots \rangle \\
\text{COM} & \{\ldots\} \\
\text{BG} & \{\ldots\} \\
\text{QUD} & \left\langle \boxed{Q}, \ldots \right\rangle \\
\text{UTT} & \left\langle \left\{ \boxed{U_2}
\begin{bmatrix}
\text{PHON} & \langle \ldots \rangle \\
\text{CONT} & P(\boxed{X})
\end{bmatrix}, \boxed{S_1}
\begin{bmatrix}
\text{PHON} & \langle \ldots \rangle \\
\text{CONT} & \boxed{Q}[\ldots \boxed{X} \ldots]
\end{bmatrix}, \boxed{U_1}\left[\text{PHON} \ \langle \ldots \rangle\right], \ldots \right\} \right\rangle \\
\text{PENDING} & \left\langle \boxed{U_2}, \boxed{U_1} \right\rangle
\end{bmatrix}
$$

Providing that it can be successfully grounded (if not, another CR will be generated and a nested clarification sequence will begin), and provides an answer to the system CR, the IS is

updated as shown in AVM (353). Grounding the new user utterance will add a new proposition to COM and a new parameter to the background BG, concerning the correct intended reference of the parameter that has been clarified. The new utterance has also been removed from the pending stack, and the CR question from QUD now that it has been answered.

$$
(353) \quad
\begin{bmatrix}
\text{AGENDA} & \langle \dots \rangle \\[4pt]
\text{COM} & \{\dots, P(\boxed{X}), \dots\} \\[4pt]
\text{BG} & \{\dots, \boxed{X}, \dots\} \\[4pt]
\text{QUD} & \langle \dots \rangle \\[4pt]
\text{UTT} & \left\langle \left\{ \boxed{U_2}\begin{bmatrix} \text{PHON} & \langle \dots \rangle \\ \text{CONT} & P(\boxed{X}) \end{bmatrix}, \boxed{S_1}\begin{bmatrix} \text{PHON} & \langle \dots \rangle \\ \text{CONT} & \boxed{Q}\big[\dots \boxed{X} \dots\big] \end{bmatrix}, \boxed{U_1}\big[\text{PHON} \ \langle \dots \rangle\big], \dots \right\} \right\rangle \\[4pt]
\text{PENDING} & \langle \boxed{U_1} \rangle
\end{bmatrix}
$$

Now, providing that the clarified parameter was the only problem with grounding the original source utterance, there is enough information in context to be able to ground it successfully. The standard update rules apply, and the originally intended effects of the utterance are carried out (in the case of the utterance asking a question Q, as shown below, Q is added to QUD and an action to answer it is added to the agenda).
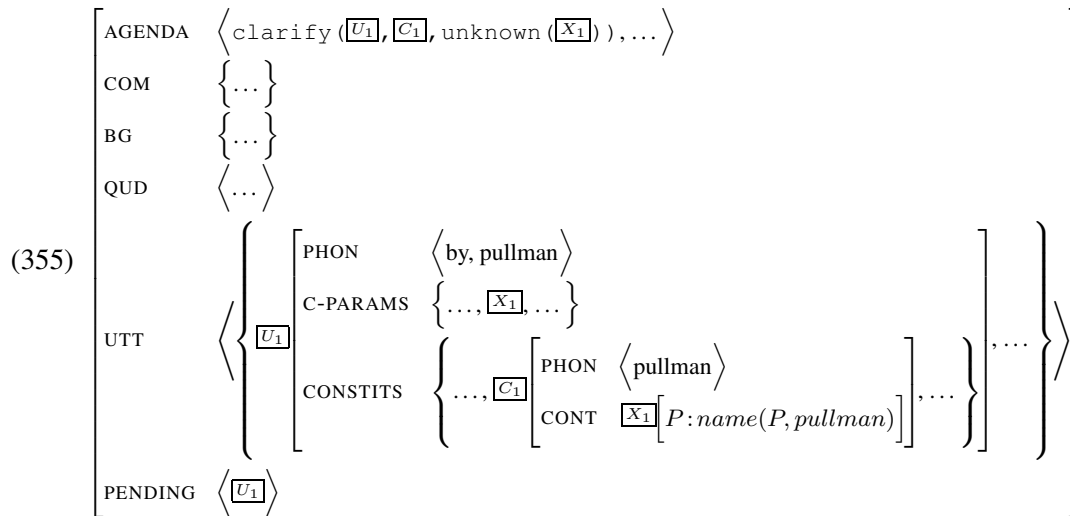
$$
(354) \quad
\begin{bmatrix}
\text{AGENDA} & \langle \texttt{respond}(Q), \dots \rangle \\[4pt]
\text{COM} & \{\dots, P(\boxed{X}), \dots\} \\[4pt]
\text{BG} & \{\dots, \boxed{X}, \dots\} \\[4pt]
\text{QUD} & \langle Q, \dots \rangle \\[4pt]
\text{UTT} & \left\langle \left\{ \boxed{U_2}\begin{bmatrix} \text{PHON} & \langle \dots \rangle \\ \text{CONT} & P(\boxed{X}) \end{bmatrix}, \boxed{S_1}\begin{bmatrix} \text{PHON} & \langle \dots \rangle \\ \text{CONT} & \boxed{Q}\big[\dots \boxed{X} \dots\big] \end{bmatrix}, \boxed{U_1}\begin{bmatrix} \text{PHON} & \langle \dots \rangle \\ \text{CONT} & ask(usr, Q) \end{bmatrix}, \dots \right\} \right\rangle \\[4pt]
\text{PENDING} & \langle \rangle
\end{bmatrix}
$$

If there are further parameters that are still problematic, application of the standard grounding rules will again fail, and the clarification update rules will produce a new `clarify` action, and thereby a new CR relevant to the next problematic parameter. The process then iterates until grounding succeeds.
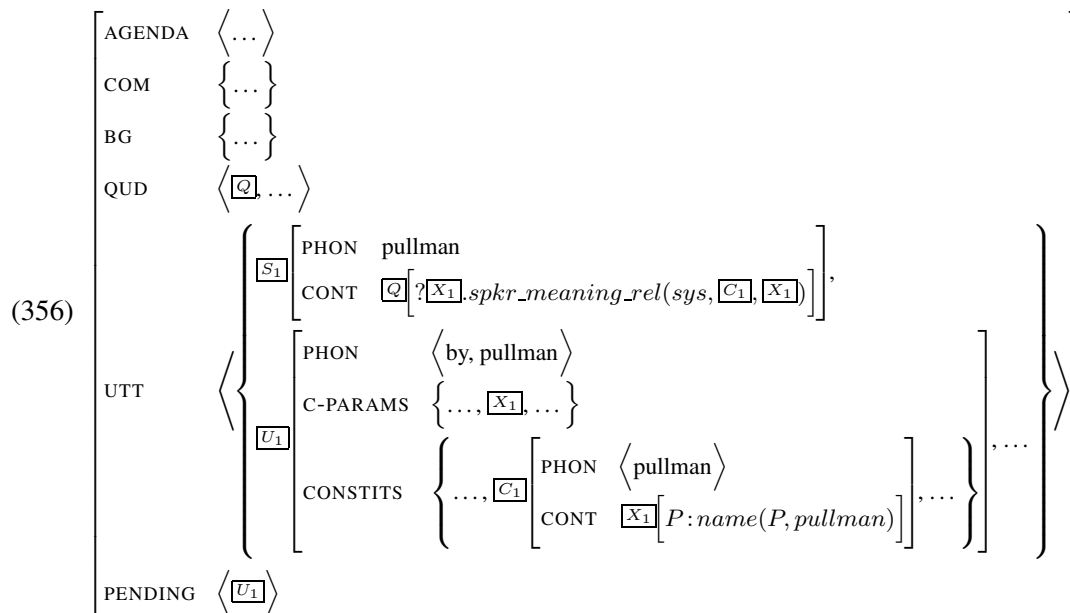
Unknown Parameter Example

In the first example *"Pullman?"*, the source utterance contains an out-of-lexicon word, and is therefore prevented from being fully grounded by the presence of a parameter whose referent cannot be identified. The clarification update rule `clarifyUnknownParameter`

therefore adds an appropriate action to clarify the parameter and associated constituent:

$$
(355) \quad
\begin{bmatrix}
\text{AGENDA} & \left\langle \texttt{clarify}(\boxed{U_1}, \boxed{C_1}, \texttt{unknown}(\boxed{X_1})), \dots \right\rangle \\
\text{COM} & \{\dots\} \\
\text{BG} & \{\dots\} \\
\text{QUD} & \langle \dots \rangle \\
\text{UTT} & \left\langle \left\{ \boxed{U_1} \begin{bmatrix} \text{PHON} & \langle \text{by, pullman} \rangle \\ \text{C-PARAMS} & \{\dots, \boxed{X_1}, \dots\} \\ \text{CONSTITS} & \left\{ \dots, \boxed{C_1} \begin{bmatrix} \text{PHON} & \langle \text{pullman} \rangle \\ \text{CONT} & \boxed{X_1}[P\!:\!name(P, pullman)] \end{bmatrix}, \dots \right\} \end{bmatrix} \right\}, \dots \right\rangle \\
\text{PENDING} & \left\langle \boxed{U_1} \right\rangle
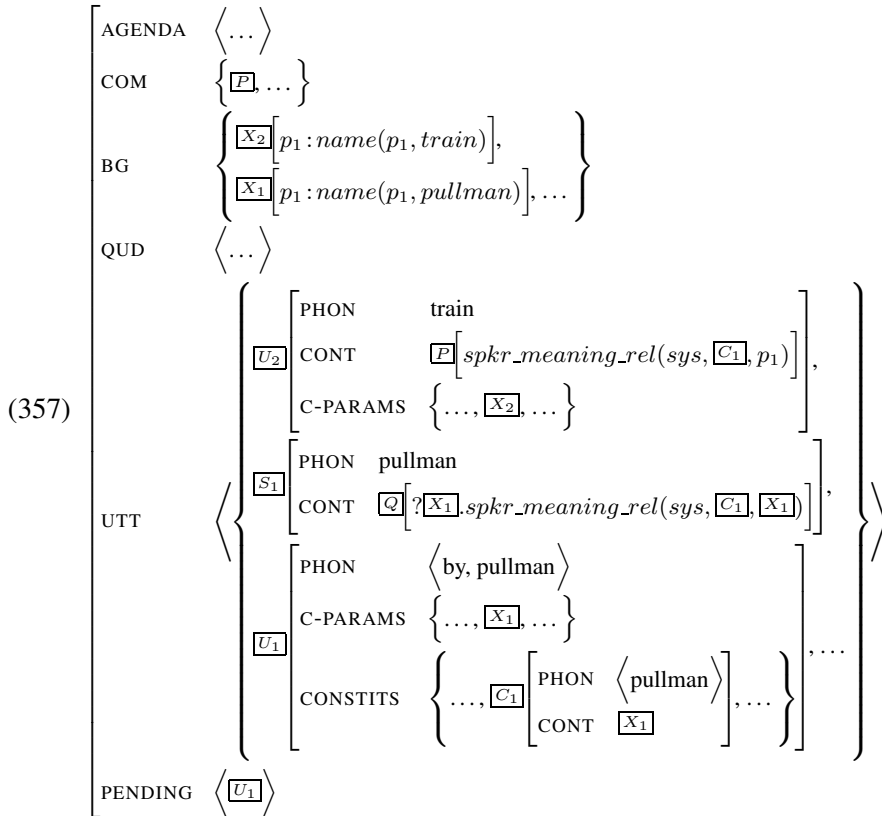\end{bmatrix}
$$

The selection rule `selectClarify` now causes a constituent CR to be generated and asked. This system utterance can of course be grounded (identifying the CP roles as usual, and an appropriate coerced MAX-QUD and SAL-UTT via the normal integration rules for CRs), and thus introduces the CR question to QUD:

$$
(356) \quad
\begin{bmatrix}
\text{AGENDA} & \langle \dots \rangle \\
\text{COM} & \{\dots\} \\
\text{BG} & \{\dots\} \\
\text{QUD} & \langle \boxed{Q}, \dots \rangle \\
\text{UTT} & \left\langle \left\{ \boxed{S_1} \begin{bmatrix} \text{PHON} & \text{pullman} \\ \text{CONT} & \boxed{Q}[?\boxed{X_1}.spkr\_meaning\_rel(sys, \boxed{C_1}, \boxed{X_1})] \end{bmatrix}, \right. \right. \\
& \left. \left. \boxed{U_1} \begin{bmatrix} \text{PHON} & \langle \text{by, pullman} \rangle \\ \text{C-PARAMS} & \{\dots, \boxed{X_1}, \dots\} \\ \text{CONSTITS} & \left\{ \dots, \boxed{C_1} \begin{bmatrix} \text{PHON} & \langle \text{pullman} \rangle \\ \text{CONT} & \boxed{X_1}[P\!:\!name(P, pullman)] \end{bmatrix}, \dots \right\} \end{bmatrix} \right\}, \dots \right\rangle \\
\text{PENDING} & \left\langle \boxed{U_1} \right\rangle
\end{bmatrix}
$$

The next user utterance provides an answer. First it must be grounded as usual: as it is elliptical, parameters for MAX-QUD and SAL-UTT must be found – this succeeds as the relevant CR question that it answers is the top member of QUD, and similarly for SAL-UTT, although this is not shown here. A referent for the predicate referred to by *train* must also be found – as this is available in the domain lexicon, this succeeds and the utterance is fully grounded (with this predicate p_1 explicitly introduced into the background). The standard

`integrateAnswer` rule applies, removing the CR question from QUD and adding a new belief to COM – that the source constituent *pullman* referred to the predicate p_1.
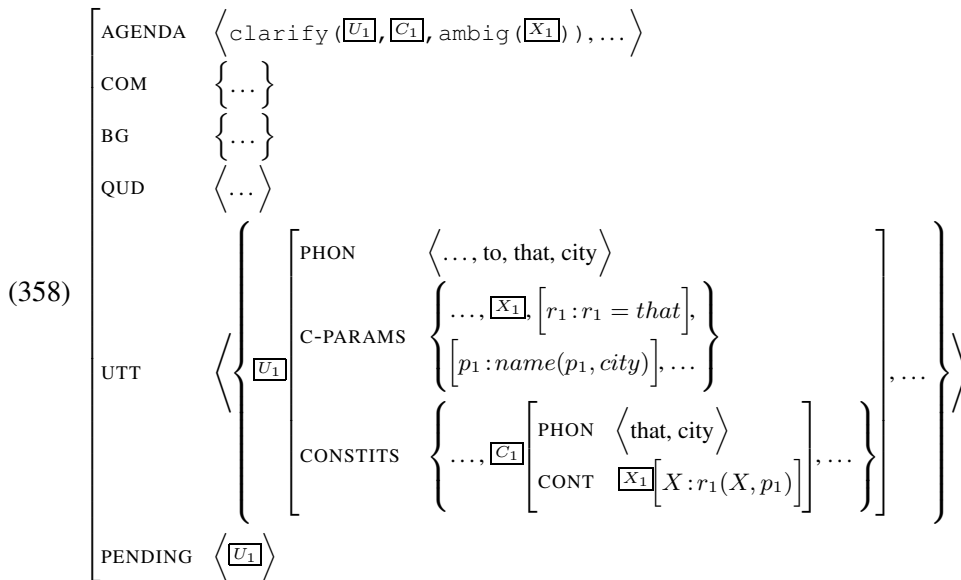
Now a IS management rule `introduceBackground` applies, which increments the background with information contained in the beliefs in COM: in this case, explicitly introducing the fact that the referent of the source constituent is p_1.

$$
(357) \quad
\begin{bmatrix}
\text{AGENDA} & \left\langle \dots \right\rangle \\[4pt]
\text{COM} & \left\{ \boxed{P}, \dots \right\} \\[4pt]
\text{BG} & \left\{
\begin{array}{l}
\boxed{X_2}\left[p_1 : name(p_1, train) \right], \\[4pt]
\boxed{X_1}\left[p_1 : name(p_1, pullman) \right], \dots
\end{array}
\right\} \\[4pt]
\text{QUD} & \left\langle \dots \right\rangle \\[4pt]
\text{UTT} & \left\langle
\left\{
\begin{array}{l}
\boxed{U_2}\begin{bmatrix}
\text{PHON} & train \\
\text{CONT} & \boxed{P}\left[spkr\_meaning\_rel(sys, \boxed{C_1}, p_1) \right], \\
\text{C-PARAMS} & \left\{ \dots, \boxed{X_2}, \dots \right\}
\end{bmatrix}, \\[14pt]
\boxed{S_1}\begin{bmatrix}
\text{PHON} & pullman \\
\text{CONT} & \boxed{Q}\left[?\boxed{X_1}.spkr\_meaning\_rel(sys, \boxed{C_1}, \boxed{X_1}) \right]
\end{bmatrix}, \\[14pt]
\boxed{U_1}\begin{bmatrix}
\text{PHON} & \left\langle by, pullman \right\rangle \\
\text{C-PARAMS} & \left\{ \dots, \boxed{X_1}, \dots \right\} \\
\text{CONSTITS} & \left\{ \dots, \boxed{C_1}\begin{bmatrix} \text{PHON} & \left\langle pullman \right\rangle \\ \text{CONT} & \boxed{X_1} \end{bmatrix}, \dots \right\}
\end{bmatrix}, \dots
\end{array}
\right\}
\right\rangle \\[20pt]
\text{PENDING} & \left\langle \boxed{U_1} \right\rangle
\end{bmatrix}
$$

Now the background contains enough information to ground the original problematic utterance, and the dialogue continues as normal. For the rest of the dialogue, the meaning of the new word *pullman* remains defined by the parameter in the IS background, and subsequent uses will therefore be able to be grounded and interpreted as normal. For this dialogue, then, the system has effectively learnt this new out-of-vocabulary word. The new word could even be added to the lexicon, but this step is not currently made.

Ambiguous Parameter Example

The second example *"Which city?"* follows the same pattern: this time the reason for grounding failure is different, and the resulting system CR generated reflects this.

(358)
$$
\begin{bmatrix}
\text{AGENDA} & \left\langle \texttt{clarify}(\boxed{U_1}, \boxed{C_1}, \texttt{ambig}(\boxed{X_1})), \ldots \right\rangle \\
\text{COM} & \{\ldots\} \\
\text{BG} & \{\ldots\} \\
\text{QUD} & \langle\ldots\rangle \\
\text{UTT} & \left\langle \boxed{U_1} \begin{bmatrix} \text{PHON} & \langle \ldots, \text{to, that, city} \rangle \\ \text{C-PARAMS} & \left\{ \ldots, \boxed{X_1}, \left[r_1 : r_1 = that \right], \left[p_1 : name(p_1, city) \right], \ldots \right\} \\ \text{CONSTITS} & \left\{ \ldots, \boxed{C_1} \begin{bmatrix} \text{PHON} & \langle \text{that, city} \rangle \\ \text{CONT} & \boxed{X_1}\left[X : r_1(X, p_1) \right] \end{bmatrix}, \ldots \right\} \end{bmatrix}, \ldots \right\rangle \\
\text{PENDING} & \left\langle \boxed{U_1} \right\rangle
\end{bmatrix}
$$

This time the grounding process determines many possible referents for X, as the domain contains many cities, none of which have been explicitly introduced into the IS background by being raised in the dialogue so far. If the BG contained exactly one parameter referring to a city (introduced by a previous utterance discussing, say, London), grounding could succeed.

The resulting clarification action causes an appropriate selection rule to be triggered: in the case of definite NPs, reprise sluice CRs are generated as these seem to be most common and intuitively seem to give the most disambiguating information due to the inclusion of the CN (see section 6.4.3). The rest of the process is as in the previous example.

Inconsistent Parameter Example

The third example is similar again, but this time the cause of grounding failure is the consistency check: the user utterance can only be grounded in such a way that it conflicts with a previously established belief in COM (that the user wants to go to Paris, as stated in the user's previous assertion). As shown below, grounding can succeed, but only by instantiating P to refer to the known predicate $p_1 = go$ and X to the known referent London (given by the domain, and introduced explicitly as a new variable x_2). This would give the content of the utterance as asserting a proposition $p_1(usr, x_2)$, which conflicts with the belief $p_1(usr, x_1)$

in COM:

(359)

$$
\begin{bmatrix}
\text{AGENDA} & \langle \texttt{clarify}(\boxed{U_1}, \boxed{C_1}, \texttt{inconsistent}(\boxed{X_1})), \dots \rangle \\[4pt]
\text{COM} & \{ p_1(usr, x_1), \dots \} \\[4pt]
\text{BG} & \left\{ \begin{array}{l} \big[p_1\!:\!name(p_1, go)\big], \big[x_1\!:\!name(x_1, paris)\big], \\ \big[x_2\!:\!name(x_2, london)\big], \dots \end{array} \right\} \\[8pt]
\text{QUD} & \langle \dots \rangle \\[4pt]
\text{UTT} & \left\langle \left\{ \boxed{U_1} \begin{bmatrix} \text{PHON} & \langle \dots, go, to, london \rangle \\ \text{CONT} & assert(usr, P(usr, X)) \\ \text{C-PARAMS} & \left\{ \dots, \boxed{X_1}, \big[P\!:\!name(P, go)\big], \dots \right\} \\ \text{CONSTITS} & \left\{ \dots, \boxed{C_1} \begin{bmatrix} \text{PHON} & \langle london \rangle \\ \text{CONT} & \boxed{X_1}\big[X\!:\!name(X, london)\big] \end{bmatrix}, \dots \right\} \end{bmatrix} \right\}, \dots \right\rangle \\[8pt]
\text{PENDING} & \langle \boxed{U_1} \rangle
\end{bmatrix}
$$

The parameter and constituent associated with X is taken to be the cause of the inconsistency, as substituting a value other than x_2 would have been consistent with the current beliefs; the problematic belief is removed from COM and the resulting CR generated is a clausal fragment querying whether the user really intended London to be referred to:

(360)

$$
\begin{bmatrix}
\text{AGENDA} & \langle \dots \rangle \\[4pt]
\text{COM} & \{ \dots \} \\[4pt]
\text{BG} & \left\{ \begin{array}{l} \big[p_1\!:\!name(p_1, go)\big], \big[x_1\!:\!name(x_1, paris)\big], \\ \big[x_2\!:\!name(x_2, london)\big], \dots \end{array} \right\} \\[8pt]
\text{QUD} & \langle \boxed{Q}, \dots \rangle \\[4pt]
\text{UTT} & \left\langle \left\{ \begin{array}{l} \boxed{S_1} \begin{bmatrix} \text{PHON} & \langle london \rangle \\ \text{CONT} & \boxed{Q}\big[?.assert(usr, p_1(usr, x_2))\big] \end{bmatrix}, \\[10pt] \boxed{U_1} \begin{bmatrix} \text{PHON} & \langle \dots, go, to, london \rangle \\ \text{CONT} & assert(usr, P(usr, X)) \\ \text{C-PARAMS} & \left\{ \dots, \boxed{X_1}, \big[P\!:\!name(P, go)\big], \dots \right\} \\ \text{CONSTITS} & \left\{ \dots, \boxed{C_1} \begin{bmatrix} \text{PHON} & \langle london \rangle \\ \text{CONT} & \boxed{X_1}\big[X\!:\!name(X, london)\big] \end{bmatrix}, \dots \right\} \end{bmatrix} \end{array} \right\}, \dots \right\rangle \\[8pt]
\text{PENDING} & \langle \boxed{U_1} \rangle
\end{bmatrix}
$$

An affirmative answer *"yes"* to this question can now be successfully grounded (the only

parameters that must be instantiated are those relating to MAX-QUD and SAL-UTT for ellipsis resolution). As the original conflicting belief has been removed, no inconsistency remains, and the answer introduces a new belief (that the destination is indeed London – a negative answer *"No, Paris"* would of course introduce a new belief identical to the original one that had been removed). The dialogue then proceeds.

6.6   Summary

The chapter has introduced the CLARIE system, which puts together the observations and analysis of the previous chapters into a dialogue system which can handle many forms of clarificational dialogue. We have seen how an approach to grounding and contextual coercion can be defined within the system's update rules, and how this and a suitable grammar can combine to handle CRs generated by both system and user.

In particular, we have seen that a system defined in this way can:

- Clarify ambiguous reference;

- Discuss and learn new out-of-vocabulary words;

- Discuss contradictory information and revise its beliefs accordingly;

- Respond to user CRs in a suitable way.

This is achieved without having to use heavyweight inference about utterances or their relation to each other, or having to model or reason about the user's beliefs or context. It also treats clarificational dialogue in the same way as standard dialogue, and CRs in the same way as standard utterances, in the way they are represented in and have effects on the IS.

Chapter 7

Conclusions

This chapter takes a look back at the main findings of the thesis in general, and at the specific conclusions of each of the four main chapters. It then examines some arising issues and areas in which this work might benefit from further investigation and perhaps lead to further insights.

7.1   Empirical Findings

The corpus studies and experiments presented in chapter 3 showed that CRs are a relatively common phenomenon in dialogue, and developed an ontology of the possible forms of CR together with their possible (and likely) readings, all of which seemed to be derivable using a few defined contextual operations rather than having to rely on general inference. It then went on to show what the possible (and likely) sources of various CRs are, how and when they are likely to be answered, and how some attempts at form and reading disambiguation can be made. Two points in particular seem worth discussing briefly here: the apparent nature of CR sources, and the nature of CRs themselves.

Sources of Clarification

Firstly, both corpus and experimental studies showed that there is a significant difference between the clarificational potential of content and function words. Function words are unlikely to get clarified at all: not only are function words rarely the source of CRs in naturally occurring dialogue, echoed function words injected into dialogues are very hard to interpret as reprise fragments.

Secondly, and less obviously to be expected, verbs also seem only rarely to be sources of CRs, while nouns and various noun phrases are very commonly clarified. However, this is not an effect of overall frequency, nor does it seem to be one of the relative ease of generating or understanding CRs: experiments suggested that echoed verbs could be interpreted as reprises

as readily as nouns could. There must be something in the nature of verbs — semantics? information content? mutual knowledge? — that makes them less likely to be clarified.

CRs vs. Standard Questions

In some ways there are significant differences between CRs and other questions. Many CRs are not answered. Those that are answered usually get answered in the very next turn, a pattern which contrasts strongly with other questions in general, although it is quite like that for yes/no questions.

However, in other ways they don't seem so different. Although they concern other utterances, they seem to ask questions that can be paraphrased in pretty straightforward ways. Like other questions, they can be answered directly or indirectly, with full sentences or fragments; yes/no versions seem to be answered more often with yes/no answers, and *wh*-versions with corresponding fragments. More significantly, perhaps, they don't *have* to be answered immediately — other questions & answers can come between them and their answers, including clarification sequences concerning the CR itself — suggesting that they behave more like standard questions (introducing questions under discussion to the context) than some other kind of special grounding acts requiring immediate repair.

7.2   Semantic Representation

Chapter 4 then investigated the meaning of reprise questions across various word and phrase types, concentrating on nouns and noun phrases as these seem to be such common CR sources. This resulted in a semantic representation whereby most (content) words and phrases denote contextually dependent predicates or individuals, rather than using higher-order representations.

Lower-Order Noun Phrases

Using reprise questions as semantic probes offers a strong criterion for assigning denotations, that they should not only combine to make up compositional sentence meanings but explain why individual constituents give their observed reprise readings. As reprises of NPs really don't seem to concern higher-order sets of sets or generalised quantifiers, but very often do seem to concern individuals (or sets of individuals), a lower-order view seems to have more explanatory power: a representation of NPs as denoting witness sets, with a definite/indefinite distinction expressed by contextual abstraction of these sets or lack thereof. This view has its complications, of course, not least for representations of relative scope and non-monotone-increasing quantifiers, but these don't seem to be insurmountable.

Semantic Inheritance

Using this strong criterion also revealed some facts about inheritance of semantics in a grammar. While reprises of NPs can concern individual referents (and so their denotations must at least contain them), reprises of their daughter determiners and nouns on their own cannot query the referents of their mothers (and so *their* denotations probably *don't* contain them). Reprises of bare verbs don't seem to be able to concern the individuals that fill the verb's argument roles, either. A grammar must not therefore assume that a NP inherits its content directly from its daughters, either just from a head daughter or by amalgamation over many daughters; nor should it assume that verbs, as heads of sentences, are associated with the entire semantic content of the sentence. This means changing some of the standard assumptions made in frameworks like HPSG.

7.3   Grammar & Ellipsis

Chapter 5 showed how a grammar can be defined which extends Ginzburg and Cooper (2004)'s approach to cover a wide range of CRs with a wide range of source types (with their various forms and readings as observed in chapter 3), and which includes the semantic representation argued for in chapter 4.

Ellipsis as Abstraction

The contextually dependent representation that has been used throughout to explain and analyse CRs, and is one of the central features of this grammar, turned out to have a useful extension to elliptical fragments. By considering utterances as abstracts, their abstracted sets can be taken to express *all* their contextual dependence, including the dependence of fragments on contextual questions to fully specify their semantic content. This allows a view of elliptical fragments as abstracts which remains close to the spirit of (Ginzburg et al., 2001a), but makes the details of interaction with context more explicit. Fragments are contextually dependent, and have to be grounded, just like other utterances – it's just that there's even more to ground (more information to find in context).

7.4   Grounding & Dialogue

The CLARIE system of chapter 6 showed how a basic dialogue system can be implemented which can handle many forms of clarificational dialogue, being able to clarify unknown or surprising reference and meaning, and allowing users to do the same. The empirical and semantic findings of chapters 3 and 4 allowed the grounding process to interpret and disambiguate user CRs in a principled way, and allowed the selection process to choose system CR

forms similarly, with these processes being smoothly integrated into an information-state-based approach to dialogue management.

Inference vs. Grounding

Importantly, this is achieved without having to use heavyweight inference about utterances or their relation to each other, or having to model or reason about the user's beliefs or context. The grammar assigns straightforward (although heavily contextually dependent, and often ambiguous) representations to CRs; and a simple grounding process then applies these contextually dependent representations to the current context. Problems with this process, and with particular contextual parameters which must be instantiated, lead to clarification.

By assuming that this process applies for both system and user, and by assuming a limited set of pragmatic operations which allow elliptical and reprise utterances to be interpreted, system CRs can be generated and user CRs interpreted without having to explicitly model or reason about the other participant.

Integration into the Dialogue

The implementation also demonstrated that CRs don't have to be treated in a significantly different way from other utterances. They can be parsed and given an interpretation by the same grammar, taken to have similar effects on an information state (raising new questions for discussion), and reacted to in a similar way (by answering elliptically or otherwise, or indeed by clarifying them if necessary). There are differences: specific pragmatic operations are required to license their elliptical reprise forms, both in generation and interpretation; and finding their answers must involve looking into an utterance record; but there is no need to treat them as having a fundamentally different character, denoting a different type of object or move, or needing to be processed by a separate module.

7.5   Summary

In summary, then, CRs may have some idiosyncratic properties, but it seems reasonable to treat them pretty much as normal questions, and to treat clarificational dialogue pretty much as normal dialogue. CRs ask about other utterances, yes, but this is simpler than it might first appear: they ask about utterances' identity or intended meaning (and *that* is simpler than it might first appear too). While it may well be the case that some CRs, or some answers to CRs, might require some serious inference to understand, we should not forget that the same applies for normal questions too: and what we have seen suggests that we don't need it most of the time, but that we can get a long way by combining a suitable grammatical analysis with a simple grounding process and a simple set of dialogue processing rules.

7.6 Arising Issues and Further Work

Practical Issues

Implementation   The implementational part of this thesis obviously leaves much to be desired as far as a practical, usable system is concerned. The CLARIE system as described in chapter 6 is only a prototype and could benefit from extension in many ways: a larger lexicon, a grammar with a wider coverage, a more realistic domain, a speech interface. Given that some of these are present in versions of GoDiS and IBiS, transfer of modules from one system to another should be feasible in most cases.

Speech   Adding speech recognition will be challenging – particularly concerning the interaction of a standard recogniser with a treatment of unknown words such as outlined here. It is the business of a standard speech recogniser, of course, to make a best guess at a known word, rather than hypothesise about unknown ones. However, there is good reason to believe that a dual-recogniser approach (Hockey et al., 2002; Gorrell et al., 2002; Dusan and Flanagan, 2002) could fit in here (see Gorrell, 2003, in particular). There are other interesting issues too, for example using low speech recognition confidence scores to prompt lexical CRs (see Larsson, 2002).

Intonation   Adding a speech recognition interface would also open up the possibility of using intonation, particularly pitch contours, to help disambiguate CRs from other fragments, and to disambiguate the elliptical forms and readings of CRs from one another. There are certainly indications that reprise questions can be distinguished from statements (Srinivasan and Massaro, 2003), and it seems quite possible that more can be done (particularly distinguishing reprise gaps from reprise fragments).

Ellipsis & Disambiguation   Little attention has been paid here to disambiguating fragments between CRs and non-CRs, or of course between non-CR fragments of various types. In order to extend coverage and move towards a practical system, a grammar and treatment of ellipsis will be required that can treat (and disambiguate between) fragments of all types, CR and non-CR. Similarly little has been paid to identifying the source of a CR where more than one potential source is present in the most recent utterance. Some initial steps have been made into classifying and identifying fragment types by (Fernández and Ginzburg, 2002), and into identifying antecedents by (Fernández et al., 2004b, forthcoming), and integrating that approach with the CR approach here might be very worthwhile. It seems likely, though, that once many sources of information are included (not only the PoS category and recency that have been considered here, but antecedent syntax, semantics and perhaps intonation) a more complex approach than a simple ordered set of grounding rules may be required –

for example, a probabilistic or machine-learning-based approach (decision tree or rule-based learners might provide more complex but still interpretable rule sets; alternatively, maximum entropy methods might be particularly well-suited to combining such different information sources).

Confidence Levels   Use of a probabilistic approach would also allow the issue of probability thresholds and confidence levels to be taken up. It seems quite possible that the approach outlined here could be combined with the confidence score approach of e.g. Gabsdil and Bos (2003) to give the benefits of both, and allow combination with explicit and implicit confirmation behaviour to give a more realistic and useful system.

Theoretical Issues

Givenness & Presupposition   The contextual abstraction approach has allowed a representation that explicitly requires grounding of elements that must be given in context (referents of names and definites, as well as contextual questions and utterances for fragments and CRs). Can this be extended to other givenness phenomena? It might well be possible to approach topic/focus distinctions too (via the same sort of question-under-discussion treatment already proposed for fragments), and possibly even presupposition – in sentences such as *"Has Bo stopped smoking?"*, perhaps *stopped* could be said to add a contextual parameter which must be grounded by finding an event or state of Bo smoking or starting to smoke. Can a reprise *"Stopped?"* mean something like *"But when did Bo **start** smoking?"*. Perhaps it can. If so, there is potential here.

Utterance Plans   As chapter 4 discussed, most CRs (especially reprises) seem to concern semantic content or word/phrase identity. But some seem to query the relevance of the utterance to the discourse, or perhaps the speaker's intentions or the plan behind making the utterance. In these cases, answering them in a suitable manner does not seem quite as straightforward. What is really being asked for? How is it best expressed? More investigation is needed for these kind of questions, and it might be useful not only purely in terms of allowing these CRs to be processed, but in terms of establishing what really goes to make up speaker's intentions or utterance plans, and how we can get at them.

An Alternative Approach?   Utterances have been represented throughout as *simultaneous* λ-abstracts. This has served very well as a general representation of their contextual dependence at the utterance level. However, it doesn't seem to fit well with incrementality in processing – the fact that humans do process sentences and resolve at least some references in a left-to-right fashion rather than waiting until the end of a sentence. The fact that grounding of contextual parameters is independent of any existentially quantified elements being intro-

duced has also complicated the account of anaphora and scope somewhat: introduction of the B-PARAMS feature is required to account for definites (including the arguments of functional NPs) which bind intrasententially. A neater approach might therefore be one in which parameters are not just simultaneously abstracted or quantified, but made part of an ordered process which integrates the grounding of dependent referents *in* context (for definites) and the addition of new referents *to* context (for indefinites), thus incorporating some of the insights of dynamic semantics. Some initial steps towards this have been made in (Purver and Fernández, 2003).

Bibliography

Abbott, Barbara (2003). Definiteness and indefiniteness. In Horn, L. and Ward, G., editors, *Handbook of Pragmatics*. Blackwell.

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

Allwood, Jens (2000). An activity based approach to pragmatics. In Bunt, H. and Black, W., editors, *Abduction, Belief and Context in Dialogue: Studies in Computational Pragmatics*, pages 47–80. John Benjamins. Also published as Gothenburg Papers in Theoretical Linguistics 76, 1992.

Allwood, Jens, Nivre, Joakim, and Ahlsén, Elisabeth (1992). On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9(1):1–26. Also published as Gothenburg Papers in Theoretical Linguistics 64.

Anderson, Anne, Bader, Miles, Bard, Ellen, Boyle, Elizabeth, Doherty, Gwyneth, Garrod, Simon, Isard, Stephen, Kowtko, Jacqueline, McAllister, Jan, Miller, Jim, Sotillo, Catherine, Thompson, Henry, and Weinert, Regina (1991). The HCRC map task data. *Language and Speech*, 34(4):351–366.

Artstein, Ron (2002). A focus semantics for echo questions. In Bende-Farkas, Á. and Riester, A., editors, *Proceedings of the Workshop on Information Structure in Context*, pages 98–107, Stuttgart. IMS.

Asher, Nicholas (1993). *Reference to Abstract Objects in Discourse*, volume 50 of *Studies in Linguistics and Philosophy*. Kluwer Academic Publishers.

Asher, Nicholas and Lascarides, Alex (2003). *Logics of Conversation*. Cambridge University Press.

Asudeh, Ash and Crouch, Richard (2002). Glue semantics for HPSG. In van Eynde, F., Hellan, L., and Beermann, D., editors, *Proceedings of the HPSG '01 Conference*. CSLI Publications.

Barg, Petra and Walther, Markus (1998). Processing unknown words in HPSG. In *Proceedings of COLING-ACL'98*, volume 1, pages 91–95.

Barwise, Jon and Cooper, Robin (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4:159–219.

Barwise, Jon and Perry, John (1983). *Situations and Attitudes*. MIT Press.

Bavelas, Janet B., Chovil, Nicole, Lawrie, D., and Wade, L. (1992). Interactive gestures. *Discourse Processes*, 15:469–489.

Bernsen, Niels Ole, Dybkjaer, Laila, and Kolodnytsky, Mykola (2002). The NITE workbench: a tool for the annotation of natural interactivity and multimodal data. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC)*, pages 43–49, Las Palmas.

Blakemore, Diane (1994). Echo questions: a pragmatic account. *Lingua*, 94(4):197–211.

Brew, Chris (1992). Letting the cat out of the bag: Generation for shake-and-bake MT. In *Proceedings of COLING-92*, pages 610–616, Nantes.

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

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

Carlson, Gregory (1977). *Reference to Kinds in English*. PhD thesis, University of Massachusetts at Amherst.

Carter, David (1992). Lexical acquisition. In Alshawi, H., editor, *The Core Language Engine*, pages 217–234. MIT Press, Cambridge, MA.

Chierchia, Gennaro (1998). Reference to kinds across languages. *Natural Language Semantics*, 6(4):339–405.

Clark, Herbert H. (1992). *Arenas of Language Use*. University of Chicago Press & CSLI.

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

Cooper, Robin (1983). Quantification and syntactic theory. *Synthese Language Library*.

Cooper, Robin (1993). Towards a general semantic framework. In Cooper, R., editor, *Integrating Semantic Theories*. ILLC/Department of Philosophy, University of Amsterdam. Deliverable R2.1.A, Dyana-2.

Cooper, Robin (1995). The role of situations in generalized quantifiers. In Lappin, S., editor, *The Handbook of Contemporary Semantic Theory*. Blackwell.

Cooper, Robin and Ginzburg, Jonathan (2002). Using dependent record types in clarification ellipsis. In Bos, J., Foster, M., and Matheson, C., editors, *Proceedings of the 6th Workshop on the Semantics and Pragmatics of Dialogue (EDILOG)*, pages 45–52, Edinburgh.

Cooper, Robin, Larsson, Staffan, Poesio, Massimo, Traum, David, and Matheson, Colin (1999). Coding instructional dialogue for information states. In *Task Oriented Instructional Dialogue (TRINDI): Deliverable 1.1*. University of Gothenburg.

Copestake, Ann and Flickinger, Dan (2000). An open-source grammar development environment and broad-coverage English grammar using HPSG. In *Proceedings of the 2nd Conference on Language Resources and Evaluation (LREC-2000)*, Athens.

Copestake, Ann, Flickinger, Dan, Sag, Ivan, and Pollard, Carl (1999). Minimal recursion semantics: An introduction. Draft.

Corblin, Francis (1996). Quantification et anaphore discursive: la référence aux complémentaires. *Langages*, 123:51–74.

Dallas, Iakovos (2001). Incorporating a grammar module in a dialogue system for the travel domain. Master's thesis, King's College, London.

Dalrymple, Mary, Shieber, Stuart M., and Pereira, Fernando C. N. (1991). Ellipsis and higher-order unification. *Linguistics and Philosophy*, 14(4):399–452.

Davidson, Donald (1980). *Essays on Actions and Events*. Clarendon Press, Oxford.

Dekker, Paul (2002). A pragmatic view upon indefinites. In von Heusinger, K., Kempson, R., and Meyer-Viol, W., editors, *Proceedings of the ESSLLI-01 Workshop on Choice Functions and Natural Language Semantics*. Working Papers of the Department of Linguistics in Konstanz.

Donnellan, Keith (1966). Reference and definite descriptions. *Philosophical Review*, 77:281–304.

Dusan, Sorin and Flanagan, James (2001). Human language acquisition by computers. In *Proceedings of the International Conference on Robotics, Distance Learning and Intelligent Communication Systems*, Malta.

Dusan, Sorin and Flanagan, James (2002). Adaptive dialog based upon multimodal language acquisition. In *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces*, Pittsburgh.

Engdahl, Elisabet, Larsson, Staffan, and Ericsson, Stina (1999). Focus-ground articulation and parallelism in a dynamic model of dialogue. In *Task Oriented Instructional Dialogue (TRINDI): Deliverable 4.1*. University of Gothenburg.

Engdahl, Elisabet and Vallduví, Enric (1996). Information packaging in HPSG. In Grover, C. and Vallduví, E., editors, *Studies in HPSG*, volume 12 of *Edinburgh Working Papers in Cognitive Science*, pages 1–31. University of Edinburgh.

Erbach, Gregor (1990). Syntactic processing of unknown words. In Jorrand, P. and Sgurev, V., editors, *Artificial Intelligence IV – methodology, systems, applications*. North-Holland, Amsterdam.

Erbach, Gregor (1991). A bottom-up algorithm for parsing and generation. CLAUS report, Universität des Saarlandes, Saarbrücken.

Erbach, Gregor (1995). Prolog with features, inheritance and templates. In *Proceedings of the 7th Conference of the European Association for Computational Linguistics*, pages 180–187.

Farkas, Donka (1997). Dependent indefinites. In Corblin, F., Godard, D., and Maradin, J.-M., editors, *Empirical Issues in Formal Syntax and Semantics*, pages 243–267. Peter Lang.

Fellbaum, Christiane, editor (1998). *WordNet: An Electronic Lexical Database*. MIT Press.

Fernández, Raquel (2002). An implemented HPSG grammar for SHARDS. Technical Report TR-02-04, Department of Computer Science, King's College London.

Fernández, Raquel and Ginzburg, Jonathan (2002). Non-sentential utterances: A corpus-based study. *Traitement Automatique des Langues*, 43(2).

Fernández, Raquel, Ginzburg, Jonathan, Gregory, Howard, and Lappin, Shalom (2004a). SHARDS: Fragment resolution in dialogue. In Bunt, H. and Muskens, R., editors, *Computing Meaning*, volume 3. Kluwer Academic Publishers. To appear.

Fernández, Raquel, Ginzburg, Jonathan, and Lappin, Shalom (2004b). Classifying ellipsis in dialogue: A machine learning approach. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING)*, Geneva.

Fernández, Raquel, Ginzburg, Jonathan, and Lappin, Shalom (forthcoming). PROFILE project results. ms.

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

Fodor, Janet and Sag, Ivan (1982). Referential and quantificational indefinites. *Linguistics and Philosophy*, 5:355–398.

Gabsdil, Malte (2003). Clarification in spoken dialogue systems. In *Proceedings of the AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*, pages 28–35, Stanford.

Gabsdil, Malte and Bos, Johan (2003). Combining acoustic confidence scores with deep semantic analysis for clarification dialogues. In *Proceedings of the 5th International Workshop on Computational Semantics (IWCS-5)*, Tilburg.

Ginzburg, Jonathan (1995). Resolving questions, I. *Language and Philosophy*, 18(5):459–527.

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

Ginzburg, Jonathan (1999). Clarification in dialogue: Meaning, content and compositionality. In *Proceedings of the 3rd Workshop on the Semantics and Pragmatics of Dialogue (Amstelogue)*. University of Amsterdam.

Ginzburg, Jonathan (2001). Fragmenting meaning: Clarification ellipsis and nominal anaphora. In Bunt, H., editor, *Computing Meaning 2: Current Issues in Computational Semantics*, Studies in Linguistics and Philosophy. Kluwer Academic Publishers.

Ginzburg, Jonathan (2003). Disentangling public from private meaning. In Smith, R. and van Kuppevelt, J., editors, *Current and New Directions in Discourse & Dialogue*, pages 183–211. Kluwer Academic Publishers.

Ginzburg, Jonathan (forthcoming). *A Semantics for Interaction in Dialogue*. CSLI Publications. Draft chapters available from: `http://www.dcs.kcl.ac.uk/staff/ginzburg`.

Ginzburg, Jonathan and Cooper, Robin (2001). Resolving ellipsis in clarification. In *Proceedings of the 39th Meeting of the ACL*, pages 236–243. Association for Computational Linguistics.

Ginzburg, Jonathan and Cooper, Robin (2004). Clarification, ellipsis, and the nature of contextual updates in dialogue. *Linguistics and Philosophy*, 27(3):297–365.

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

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

Ginzburg, Jonathan, Sag, Ivan, and Purver, Matthew (2001b). Integrating conversational move types in the grammar of conversation. In Kühnlein, P., Rieser, H., and Zeevat, H., editors, *Proceedings of the 5th Workshop on Formal Semantics and Pragmatics of Dialogue (BI-DIALOG)*, pages 45–56.

Ginzburg, Jonathan, Sag, Ivan, and Purver, Matthew (2003). Integrating conversational move types in the grammar of conversation. In Kühnlein, P., Rieser, H., and Zeevat, H., editors, *Perspectives on Dialogue in the New Millennium*, volume 114 of *Pragmatics and Beyond New Series*, pages 25–42. John Benjamins.

Gorrell, Genevieve (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)*.

Gorrell, Genevieve, Lewin, Ian, and Rayner, Manny (2002). Adding intelligent help to mixed-initiative spoken dialogue systems. In *ACL-02 Companion Volume to the Proceedings of the Conference*, page 95, Philadelphia. Association for Computational Linguistics.

Granger, Richard H. (1977). FOUL-UP: A program that figures out meanings of words from context. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence (IJCAI-77)*, volume 1, pages 172–178.

Grice, Martine, Benzmüller, Ralf, Savino, Michelina, and Andreeva, Bistra (1995). The intonation of queries and checks across languages: Data from map task dialogues. In *Proc. XIII International Congress of Phonetic Sciences, Stockholm*.

Groenendijk, Jeroen and Stokhof, Martin (1984). On the semantics of questions and the pragmatics of answers. In Landman, F. and Veltman, F., editors, *Varieties of Formal Semantics*, volume 3 of *Groningen-Amsterdam Studies in Semantics (GRASS)*, pages 143–170. Foris.

Groenendijk, Jeroen and Stokhof, Martin (1991). Dynamic predicate logic. *Linguistics and Philosophy*, 14(1):39–100.

Healey, Patrick, Purver, Matthew, King, James, Ginzburg, Jonathan, and Mills, Greg (2003). Experimenting with clarification in dialogue. In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, Boston.

Heeman, Peter and Allen, James (1995). The TRAINS 93 dialogues. Trains Technical Note 94-2, Computer Science Department, University of Rochester.

Heeman, Peter and Hirst, Graeme (1995). Collaborating on referring expressions. *Computational Linguistics*, 21(3):351–382.

Heim, Irene (1982). *The Semantics of Definite and Indefinite Noun Phrases*. PhD thesis, University of Massachusetts at Amherst.

Hobbs, Jerry (1983). An improper treatment of quantification in ordinary English. In *Proceedings of the 21st Annual Meeting*, pages 57–63. Association for Computational Linguistics.

Hobbs, Jerry (1996). Monotone decreasing quantifiers. In van Deemter, K. and Peters, S., editors, *Semantic Ambiguity and Underspecification*, number 55 in CSLI Lecture Notes, pages 55–76. CSLI Publications.

Hockey, Beth Ann (1994). Echo questions, intonation and focus. In *Proceedings of the Interdisciplinary Conference on Focus and Natural Language Processing in Celebration of the 10th Anniversary of the Journal of Semantics*, Eschwege.

Hockey, Beth Ann, Dowding, John, Aist, Gregory, and Hieronymus, Jim (2002). Targeted help and dialogue about plans. In *ACL-02 Companion Volume to the Proceedings of the Conference*, pages 100–101, Philadelphia. Association for Computational Linguistics.

Hockey, Beth Ann, Rossen-Knill, Deborah, Spejewski, Beverly, Stone, Matthew, and Isard, Stephen (1997). Can you predict answers to Yes/No questions? Yes, No and Stuff. In *Proceedings of Eurospeech '97*.

Hornby, Albert S. (1974). *Oxford Advanced Learner's Dictionary of Current English*. Oxford University Press, third edition. With the assistance of Anthony P. Cowie and J. Windsor Lewis.

Ide, Nancy and Priest-Dorman, Greg (1996). The corpus encoding standard.

Iwata, Seizi (2003). Echo questions are interrogatives? another version of a metarepresentational analysis. *Linguistics and Philosophy*, 26(2):185–254.

Janda, Richard (1985). Echo-questions are evidence for what? In *Papers from the 21st Regional Meeting of the Chicago Linguistic Society*, pages 171–188. CLS.

Kamp, Hans and Reyle, Uwe (1993). *From Discourse To Logic*. Kluwer Academic Publishers.

Keenan, Edward and Stavi, Jonathan (1986). A semantic characterization of natural language determiners. *Linguistics and Philosophy*, 9:253–326.

Keenan, Edward and Westerståhl, Dag (1997). Generalized quantifiers in linguistics and logic. In van Benthem, J. and ter Meulen, A., editors, *Handbook of Logic and Language*, pages 837–893. Elsevier.

Kempson, Ruth, Meyer-Viol, Wilfried, and Gabbay, Dov (2001). *Dynamic Syntax: The Flow of Language Understanding*. Blackwell.

Kibble, Rodger (1997a). Complement anaphora and dynamic binding. In Lawson, A., editor, *Proceedings of the 7th annual conference on Semantics and Linguistic Theory (SALT VII)*. Cornell University.

Kibble, Rodger (1997b). Complement anaphora and monotonicity. In Morrill, G., Kruijff, G.-J., and Oehrle, R., editors, *Proceedings of the Formal Grammar conference*, pages 125–136.

Kilgarriff, Adam (1997). Putting frequencies in the dictionary. *International Journal of Lexicography*, 10(2):135–155.

Knight, Kevin (1996). Learning word meanings by instruction. In *Proceedings of the 13th National Conference on Artifical Intelligence*, pages 447–454. AAAI/IAAI.

Kripke, Saul (1977). Speaker's reference and semantic reference. In French, P., Uehling, T., and Wettstein, H., editors, *Perspectives in the Philosophy of Language*, number 2 in Midwest Studies in Philosophy, pages 6–27. University of Minnesota Press.

Lappin, Shalom (2002). Salience and inference in anaphora resolution. In *Proceedings of the 4th Discourse and Anaphora Resolution Colloquium*, Lisbon. Invited Talk.

Larsson, Staffan (2000). GoDiS 1.2 developers manual. Draft.

Larsson, Staffan (2002). *Issue-based Dialogue Management*. PhD thesis, Göteborg University. Also published as Gothenburg Monographs in Linguistics 21.

Larsson, Staffan, Berman, Alexander, Bos, Johan, Grönqvist, Leif, Ljunglöf, Peter, and Traum, David (1999). TrindiKit 2.0 manual. In *Task Oriented Instructional Dialogue (TRINDI): Deliverable 5.3*. University of Gothenburg.

Larsson, Staffan, Ljunglöf, Peter, Cooper, Robin, Engdahl, Elisabet, and Ericsson, Stina (2000). GoDiS - an accommodating dialogue system. In *Proceedings of ANLP/NAACL-2000 Workshop on Conversational Systems*.

Lewin, Ian and Pulman, Stephen (1995). Inference in the resolution of ellipsis. In *Proceedings of the ESCA Workshop on Spoken Dialogue Systems*, pages 53–56.

Ludlow, Peter and Neale, Stephen (1991). Indefinite descriptions: In defense of Russell. *Linguistics and Philosophy*, 14:171–202.

Ludlow, Peter and Segal, Gabriel (2004). On a unitary analysis of definite and indefinite descriptions. In Reimer, M. and Bezuidenhout, A., editors, *Descriptions and Beyond: An Interdisciplinary Collection of Essays on Definite and Indefinite Descriptions*. Oxford University Press.

Macura, Zoran (2002). Classifying answers. BSc dissertation, King's College, London.

Mikheev, Andrei (1997). Automatic rule induction for unknown word guessing. *Computational Linguistics*, 23(3):405–423.

Montague, Richard (1974). The proper treatment of quantification in ordinary English. In Thomason, R., editor, *Formal Philosophy: Selected Papers of Richard Montague*, pages 247–270. Yale University Press.

Moore, Johanna (1993). What makes human explanations effective? In *Proceedings of the 15th Annual Meeting of the Cognitive Science Society*. Lawrence Erlbaum Associates.

Moxey, Linda and Sanford, Anthony (1987). Quantifiers and focus. *Journal of Semantics*, 5:189–206.

Moxey, Linda and Sanford, Anthony (1993). *Communicating Quantities: a Psychological Perspective*. Lawrence Erlbaum Associates.

Müller, Christoph and Strube, Michael (2003). Multi-level annotation in MMAX. In *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue*, pages 198–207, Sapporo. Association for Computational Linguistics.

Neumann, Günter (1994). *A Uniform Computational Model for Natural Language Parsing and Generation*. PhD thesis, Universität des Saarlandes, Saarbrücken.

Noh, Eun-Ju (1998). Echo questions: Metarepresentation and pragmatic enrichment. *Linguistics and Philosophy*, 21(6):603–628.

Noh, Eun-Ju (2001). *Metarepresentation: A Relevance-Theory Approach*. John Benjamins.

Nouwen, Rick (2003). Complement anaphora and interpretation. *Journal of Semantics*, 20(1):73–113.

Pedersen, Ted (1995). Automatic acquisition of noun and verb meanings. Technical Report 95-CSE-10, Southern Methodist University.

Pelletier, Francis J. (2003). Context dependence and compositionality. *Mind and Language*, 18(2):148–161.

Pickering, Martin and Garrod, Simon (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, forthcoming.

Poesio, Massimo (1993). A situation-theoretic formalization of definite description interpretation in plan elaboration dialogues. In Aczel, P., Israel, D., Katagiri, Y., and Peters, S., editors, *Situation Theory and its Applications*, volume 3, pages 339–374. CSLI Publications.

Poesio, Massimo (1994). Weak definites. In *Proceedings of the 4th Conference on Semantics and Linguistic Theory (SALT-4)*.

Poesio, Massimo and Traum, David (1997). Conversational actions and discourse situations. *Computational Intelligence*, 13(3).

Poesio, Massimo and Traum, David (1998). Towards an axiomatization of dialogue acts. In Hulstijn, J. and Nijholt, A., editors, *Proceedings of the 2nd Workshop on Formal Semantics and Pragmatics of Dialogue (Twendial)*, pages 207–222, Enschede.

Pollard, Carl and Sag, Ivan (1994). *Head Driven Phrase Structure Grammar*. University of Chicago Press and CSLI, Chicago.

Porter, Martin (1980). An algorithm for suffix stripping. *Program*, 14(3):130–137.

Prince, Ellen (1992). The ZPG letter: Subjects, definiteness and information status. In Mann, W. and Thompson, S., editors, *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text*, pages 295–326. John Benjamins.

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

Purver, Matthew (2002). Processing unknown words in a dialogue system. In *Proceedings of the 3rd SIGdial Workshop on Discourse and Dialogue*, pages 174–183, Philadelphia. Association for Computational Linguistics.

Purver, Matthew and Fernández, Raquel (2003). Utterances as update instructions. In *Proceedings of the 7th Workshop on the Semantics and Pragmatics of Dialogue (DiaBruck)*, pages 115–122, Saarbrücken.

Purver, Matthew and Ginzburg, Jonathan (2003). Clarifying noun phrase semantics in HPSG. In Müller, S., editor, *Proceedings of the 10th International Conference on Head-Driven Phrase Structure Grammar (HPSG-03)*, pages 338–358, East Lansing. Michigan State University.

Purver, Matthew and Ginzburg, Jonathan (2004). Clarifying noun phrase semantics. *Journal of Semantics*, 21(3):283–339.

Purver, Matthew, Ginzburg, Jonathan, and Healey, Patrick (2001). On the means for clarification in dialogue. In *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue*, pages 116–125, Aalborg. Association for Computational Linguistics.

Purver, Matthew, Ginzburg, Jonathan, and Healey, Patrick (2003a). On the means for clarification in dialogue. In Smith, R. and van Kuppevelt, J., editors, *Current and New Directions in Discourse & Dialogue*, pages 235–255. Kluwer Academic Publishers.

Purver, Matthew, Healey, Patrick, King, James, Ginzburg, Jonathan, and Mills, Greg (2003b). Answering clarification questions. In *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue*, pages 23–33, Sapporo. Association for Computational Linguistics.

Ratnaparkhi, Adwait (1996). A maximum entropy part-of-speech tagger. In *Proceedings of the Empirical Methods in Natural Language Processing Conference*. University of Pennsylvania.

Reeves, Byron and Nass, Clifford (1996). *The Media Equation: How People Treat Computers, Television and New Media like Real People and Places*. Cambridge University Press.

Reinhart, Tanya (1997). Quantifier scope: How labour is divided between QR and choice functions. *Linguistics and Philosophy*, 20:335–397.

Roberts, Craige (2002). Demonstratives as definites. In van Deemter, K. and Kibble, R., editors, *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*, pages 89–136. CSLI Publications.

Rosé, Carolyn (1997). *Robust Interactive Dialogue Interpretation*. PhD thesis, School of Computer Science, Carnegie Mellon University.

Ross, John R. (1969). Guess who? In Binnick, R. I., Davison, A., Green, G., and Morgan, J., editors, *Papers from the 5th Regional Meeting of the Chicago Linguistic Society*, pages 252–286. CLS, University of Chicago.

Russell, Bertrand (1905). On denoting. *Mind*, 14:479–493.

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

Sacks, Harvey (1992). *Lectures on Conversation*. Blackwell.

Sag, Ivan and Wasow, Thomas (1999). *Syntactic Theory: A Formal Introduction*. CSLI Publications.

San-Segundo, Ruben, Montero, Juan M., Guitiérrez, Juana M., Gallardo, Ascension, Romeral, Jose D., and Pardo, Jose M. (2001). A telephone-based railway information system for spanish: Development of a methodology for spoken dialogue design. In *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue*, pages 140–148, Aalborg. Association for Computational Linguistics.

Schegloff, Emmanuel A. (1987). Some sources of misunderstanding in talk-in-interaction. *Linguistics*, 25:201–218.

Schlangen, David (2003). *A Coherence-Based Approach to the Interpretation of Non-Sentential Utterances in Dialogue*. PhD thesis, University of Edinburgh.

Schlangen, David (2004). Causes and strategies for requesting clarification in dialogue. In *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue*, Boston. Association for Computational Linguistics.

Schlangen, David and Lascarides, Alex (2002). Resolving fragments using discourse information. In Bos, J., Foster, M., and Matheson, C., editors, *Proceedings of the 6th Workshop on the Semantics and Pragmatics of Dialogue (EDILOG)*, pages 161–168, Edinburgh.

Schlangen, David and Lascarides, Alex (2003). A compositional and constraint-based approach to non-sentential utterances. In Müller, S., editor, *Proceedings of the 10th International Conference on Head-Driven Phrase Structure Grammar*, East Lansing. Michigan State University.

Schlangen, David, Lascarides, Alex, and Copestake, Ann (2003). Resolving underspecification using discourse information. In Kühnlein, P., Rieser, H., and Zeevat, H., editors, *Perspectives on Dialogue in the New Millennium*, volume 114 of *Pragmatics and Beyond New Series*, pages 287–305. John Benjamins.

Shieber, Stuart (1988). A uniform architecture for parsing and generation. In *Proceedings of the 12th International Conference on Computational Linguistics (COLING)*, pages 614–619.

Shieber, Stuart, van Noord, Gertjan, Moore, Robert, and Pereira, Fernando (1990). A semantic head-driven generation algorithm for unification-based formalisms. *Computational Linguistics*, 16(1).

Sperber, Dan and Wilson, Deirdre (1986). *Relevance: Communication and Cognition*. Blackwell.

Srinivasan, Ravindra and Massaro, Dominic (2003). Perceiving prosody from the face and voice: Distinguishing statements from echoic questions in English. *Language and Speech*, 46(1):1–22.

Strawson, Peter (1950). On referring. *Mind*, 59:320–344.

Szabolcsi, Anna (1997). Strategies for scope taking. In Szabolcsi, A., editor, *Ways of Scope Taking*, pages 109–155. Kluwer Academic Publishers.

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

Traum, David (2003). Semantics and pragmatics of questions and answers for dialogue agents. In *Proceedings of the International Workshop on Computational Semantics*, pages 380–394.

Traum, David, Fleischman, Michael, and Hovy, Eduard (2003). NL generation for virtual humans in a complex social environment. In *Papers from the AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*, pages 151–158.

van den Berg, Martin (1996). Dynamic generalised quantifiers. In van der Does, J. and van Eijck, J., editors, *Quantifiers, Logic and Language*, Studies in Linguistics and Philosophy, pages 63–94. CSLI Publications.

van der Does, Jaap and van Eijck, Jan (1996). Basic quantifier theory. In van der Does, J. and van Eijck, J., editors, *Quantifiers, Logic and Language*, pages 1–45. CSLI Publications.

van der Sandt, Rob (1992). Presupposition projection as anaphora resolution. *Journal of Semantics*, 9:333–377.

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

van Noord, Gertjan, Bouma, Gosse, Koeling, Rob, and Nederhof, Mark-Jan (1999). Robust grammatical analysis for spoken dialogue systems. *Natural Language Engineering*, 5(1):45–93.

van Rooy, Robert (2000). The specificity of indefinites. In *Proceedings of the 1998 Budapest Workshop on Indefinites*.

von Heusinger, Klaus (2000). The reference of indefinites. In von Heusinger, K. and Egli, U., editors, *Reference and Anaphoric Relations*, number 72 in Studies in Linguistics and Philosophy, pages 247–265. Kluwer Academic Publishers.

von Heusinger, Klaus (2002). Specificity and definiteness in sentence and discourse structure. *Journal of Semantics*, 19(3):245–274. Special Issue on Specificity.

Wahlster, Wolfgang, Marburger, Heinz, Jameson, Anthony, and Busemann, Stephan (1983). Over-answering yes-no questions: Extended reponses in a natural language interface to a vision system. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence (IJCAI-83)*, pages 643–646.

Wilkinson, Karina (1991). *Studies in the Semantics of Generic Noun Phrases*. PhD thesis, University of Massachusetts at Amherst.

Zernik, Uri (1987). Language acquisition: Learning a hierarchy of phrases. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence (IJCAI-87)*, volume 1, pages 125–132.

Zimmerman, Thomas E. (1999). Remarks on the epistemic rôle of discourse referents. In Moss, L., Ginzburg, J., and de Rijke, M., editors, *Logic, Language and Computation: Volume 2*, number 96 in CSLI Lecture Notes, pages 346–368. CSLI Publications.