Towards a Robust Grammar-based Dialogue System

Matthew Purver
Natural Language Processing Group
Department of Computer Science
King's College London

July 5, 2001

Abstract

Grammar-based dialogue systems provide many advantages over keyword-based approaches, not least a degree of domain independence. Systems that use grammars to build semantic-pragmatic interpretations of utterances and use these to maintain and update a dialogue information state are becoming widespread, at least in the research domain. A disadvantage of such systems is their lack of robustness in the face of imperfect communication between user and system.

In this paper I describe how a system designed to handle imperfect communication phenomena such as unknown words and clarifications will require utterance interpretations and information state to have some degree of underspecification and to contain levels of information other than semantics. I propose some possible approaches and describe a partial implementation, together with some issues associated with the development of a realistic system.

Contents

Chapter 1

Introduction

1.1 Dialogue Systems

Significant research effort has been put into producing automated dialogue systems. Such systems could be used in a wide range of applications, and may become increasingly useful (and commercially viable) with today's widespread use of mobile phones. While most telephone users have become increasingly accustomed to dealing with automated systems, these are not usually capable of *dialogue* but are programmed to elicit and store information in a predetermined fashion. Interactivity from the user's point of view is limited to providing all and only the information requested, and anything else is outside the capabilities of the system – as in the familiar style of example (1).

(1)[1]
System: Welcome to London Electricity. [...]
If you are a pre-payment, keymeter or power key customer, please press 1.
If you have a power failure, press 2.
For all account enquiries, press 3.
If you are moving home or would like to make an appointment, press 4.
For all other services hold the line for assistance.

These finite-state systems can be extended with the addition of speech recognition and generation technology to produce voice-activated versions. While these systems can deal reasonably well in domains in which the information being given or requested by the user is (a) simple and (b) predictable, other more complex domains require more advanced treatment. Even in the simple domains, they can be inflexible and time-consuming.

An wide- or open-domain system presents several challenges. The dialogue management must be advanced enough to handle a negotiative dialogue, and

[1] London Electricity enquiry line, 0800 096 9000, 25th May 2001.

must be accommodating in the sense of Larsson et al. (2000). A significant linguistic capability will be required in order to interpret and produce utterances whose domain is unknown. Above all perhaps, the aforementioned two requirements produce a third: these advanced features must be robust in the face of a number of "noisy" phenomena (unknown words, ungrammatical input, user error or self-correction, and comprehension problems on the part of the user being but a few). The research outlined in this document is primarily concerned with possible linguistic capability, and methods for providing this capability in a robust manner.

1.2 Linguistic Processing of Dialogue

In order for a dialogue system to be of a wide- or open-domain nature, some linguistic processing must be required. In an open-domain situation, one cannot define a set of keywords or phrases pertinent to the (possibly infinite) set of actions that could be performed by the user. Instead, one must somehow attach an interpretation to utterances based on their form and constituent words. In order for a dialogue system to either be truly open-domain, or to be easily applicable to many possibly complex domains, such interpretations are required. While the actions performed by the system must necessarily remain domain-dependent, the interface framework need not change.

van Noord et al. (1998) have shown that it is possible to use (domain-specific) HPSG-like grammatical rules to produce interpretations in a dialogue system. Their approach is made robust by allowing complete sub-segments of an utterance to be treated where the whole utterance cannot be processed.

1.3 Existing Dialogue Systems

A good overview of existing systems is given by Bohlin (Ljunglöf) as part of the TRINDI[2] project. They examined four dialogue systems, all of which were restricted to a particular domain (e.g. route planning), and which had a range of linguistic capability from keyword slot-filling to sophisticated grammatical analysis with pronoun and ellipsis resolution.

Recent development systems have begun to use grammar-based approaches widely (see e.g. Lemon et al., 2001). The grammatical content tends to be domain-specific and to associate conversational moves with content on a pre-determined basis (see (Ludwig, 2001) for an exception – conversational moves and illocutionary force are computed by inference, although still on a domain-specific basis).

[2] Task Oriented Instructional Dialogue. TRINDI was a joint project involving several institutions which focussed on research into human-computer dialogue systems that "enable the human to make choices in the performance of a certain task" – see `http://www.ling.gu.se/projects/trindi/`

One system that resulted from the TRINDI project is GoDiS[3] (see Larsson et al., 2000). GoDiS has no sophisticated linguistic capability, but uses domain-specific keyword/phrase spotting techniques which associate conversational moves directly with heywords. However, its dialogue handling is sophisticated: it is capable of accommodation of user utterances that contain information that could be relevant to parts of dialogue plans not currently under discussion, and capable of negotiation of specific issues. It is also designed to be easily ported between domains and as a result is modular, with many modules being domain-independent.

The starting point of the work described here was to integrate an existing simple HPSG grammar, together with a set of ellipsis resolution routines, SHARDS (see below), with an existing dialogue system, GoDiS. In this way the benefits of GoDiS's dialogue handling could be integrated with the benefits of a grammar-based interface, along with advanced handling of ellipsis. Conversely, SHARDS' handling of ellipsis (including disambiguation) could benefit from the maintaining of a full record of dialogue information state. From this starting point, further work could then investigate more complex requirements as described below.

1.4 Imperfect Communication

One assumption all the dialogue systems mentioned above have in common is that communication is perfect. While parsers are made robust to mitigate any effects of ungrammatical input, new words, noise, ambiguity etc., a resulting successful parse is taken to be a complete representation of the intended input. An unsuccessful attempt to parse results in no representation (usually followed by a system response along the lines of *"I don't understand. Please repeat."*).

This does not seem to be a very good model of human dialogue. We often cope with ungrammatical input with no problem, and construct partial representations in the face of unknown words or noise. Indeed, we would not be able to learn language if no representation could be constructed for sentences containing new words. If a dialogue system could build such partial representations, it would not only be robust (in the partial parse sense) but could then deduce the full meaning, request clarification of specific words as necessary, or even ignore the uncertainty if it is not relevant to the dialogue.

In addition, the possibility of imperfect understanding on the part of the user must be allowed for. If users request clarification of system utterances, these requests must be recognised as such (rather than being, say, misread as responses which answer the system utterances).

One of the main areas intended for investigation is therefore the building and representation of underspecified and ambiguous utterances, and the process of clarification of these utterances where required.

[3] A demonstration version of GoDiS is available at `http://www.ling.gu.se/~peb/dme-demo/frame/`

1.5 Levels of Memory

Another assumption in common is that only the content of utterances is important. The systems described above all associate utterances with some sort of logical content (usually a conversational move with associated message) and this is all that is preserved from the utterance. When taken along with the assumption of perfect communication, this seems reasonable, but once we start to consider phenomena like requests for clarification it is apparent that more information is required: at least, information about surface form (syntax and phonological form) of the utterance.

Other evidence points towards surface form being retained by humans in dialogue. For instance, echoing of syntactic forms between conversational participants (CPs) has been demonstrated by Branigan et al. (2000). Garrod and Pickering (2001) argue that CPs align their processes with each other at all levels, from conceptual to phonological.

However, 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).

Another major area for investigation is therefore the retention of information at various levels.

In this document I describe the areas introduced above in more detail, describing the proposed HPSG dialogue system in chapter 2, then dealing with clarifications in chapter 3, robust semantic processing in chapter 4, and disambiguation of utterance content in chapter 5. In each chapter I describe the work performed so far, and give a description of future research proposed. A draft schedule for this work is given in appendix A.

Implementation of the methods and ideas discussed is vital both to prove their viability and test their relative worth. It is intended that all methods will be used to extend an existing dialogue system and that a working demonstrator version will form part of the resulting thesis. Wherever possible I relate the work described to this implementation.

Chapter 2

Integrating GoDiS and SHARDS

2.1 Background

The starting point of the implementation is GoDiS (Larsson et al., 2000), described in section 1.3 above. The modular nature of GoDiS and TrindiKit enables changes to particular aspects of the system to be made easily. The first challenge was to change the input interface from a domain-specific keyword-spotting method to one based on linguistic analysis within HPSG, together with the ellipsis resolution capability of SHARDS (see Ginzburg et al., 2001a).

Further work will use this baseline system to investigate approaches to clarifications and underspecification, eventually integrating these notions to develop a robust dialogue model which includes and uses representations of surface forms.

In this chapter I describe the integration work performed so far and outline the next steps required to complete and improve the baseline system. Descriptions of work specifically directed towards clarifications and underspecification are given in later chapters.

2.2 Work to Date

The modular nature of GoDiS also allows domain and language to be changed easily, and as it is programmed in Prolog this allowed integration with SHARDS (which uses ProFIT (Erbach, 1995) on top of Prolog). All work to date has been in English and within GoDiS's "travel" domain (a simulated travel agent which books trips for the user).

2.2.1 GoDiS Interface

The interface provided with GoDiS works by spotting keywords and interpreting them in a domain-specific manner. The files `lexicon-travel-english.pl` and domain-travel.pl together specify the possible keywords and their interpretations.

```
input_form( [to|S], answer(to(C)) ) :- lexsem( S, C ), location( C ).
input_form( [in|S], answer(month(C)) ) :- lexsem( S, C), month( C ).

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

Input processing is performed as follows. The user input is converted to a list of words (represented as plain strings) by input_simpletext.pl. The interpreter interpret_simple1.pl then attempts to match the beginning of this list to its known possible keywords and phrases. If successful, the interpretation assigned to the matched phrase (a conversational move) is put on a list of moves, and the phrase is removed from the input word list: the process is then repeated with the new shorter list. If unsuccessful, the first word is removed from the list and the process is retried.

The resulting interpretation is therefore a list of conversational moves:

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

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

2.2.2 HPSG/SHARDS Interface

The interpreter module was replaced with interpret_hpsg.pl which performs a HPSG-based linguistic analysis using SHARDS. The input word list is passed to a bottom-up left-to-right parser which uses an HPSG grammar and returns a HPSG sign. As in the normal use of SHARDS, this is then passed to an ellipsis-resolution module which uses dialogue context to resolve short answers, polar answers and sluices, producing a fully interpreted HPSG sign.

2.2.3 Conversational Move Type

In order to integrate the HPSG grammar directly with the dialogue system, some method of associating utterances with conversational moves was required. A typical utterance produced by SHARDS has as its content the logical proposition (or question etc.) associated with the utterance – see AVM [1].

$$\text{[1]}\quad \left[\begin{array}{ll} \textsc{phon} & \langle \text{does,john,like,mary}\rangle \\ \textsc{cat} & \left[\textit{interrogative}\right] \\ \textsc{cont} & \left[\begin{array}{ll} \textsc{quants} & \langle\ \rangle \\ \textsc{nucl} & \left[\begin{array}{ll} \multicolumn{2}{l}{\textit{like-rel}} \\ \textsc{liker} & \boxed{1} \\ \textsc{liked} & \boxed{2}\end{array}\right] \\ \textsc{rest} & \left\langle \left[\begin{array}{ll}\multicolumn{2}{l}{\textit{name-rel}}\\ \textsc{name} & \text{john}\\ \textsc{named} & \boxed{1}\end{array}\right], \left[\begin{array}{ll}\multicolumn{2}{l}{\textit{name-rel}}\\ \textsc{name} & \text{mary}\\ \textsc{named} & \boxed{2}\end{array}\right]\right\rangle \end{array}\right]\end{array}\right]$$

The grammar within SHARDS was therefore updated to include conversational move type (CMT) along the lines proposed in (Ginzburg et al., 2001b), so that the semantic content (the value of the CONT attribute) includes the basic illocutionary force of the utterance (its basic conversational move). The move type is determined by clausal type (declarative sentences are treated as moves of type `assert`, interrogatives as type `ask`, etc. as shown in AVM [2]) or stored in the lexicon in the case of conventional examples such as greetings.

$$\text{[2]}\quad \left[\begin{array}{ll} \textsc{phon} & \langle \text{does,john,like,mary}\rangle \\ \textsc{cat} & \left[\textit{interrogative}\right] \\ \textsc{cont} & \left[\begin{array}{ll} \textsc{quants} & \langle\ \rangle \\ \textsc{nucl} & \left[\begin{array}{ll} \multicolumn{2}{l}{\textit{ask-rel}} \\ \textsc{utt} & \boxed{3} \\ \textsc{add} & \boxed{4} \\ \textsc{msg-arg}\ |\ \textsc{nucl} & \left[\begin{array}{ll} \multicolumn{2}{l}{\textit{like-rel}} \\ \textsc{liker} & \boxed{1} \\ \textsc{liked} & \boxed{2}\end{array}\right]\end{array}\right] \\ \textsc{rest} & \left\langle \left[\begin{array}{ll}\multicolumn{2}{l}{\textit{name-rel}}\\ \textsc{name} & \text{john}\\ \textsc{named} & \boxed{1}\end{array}\right], \left[\begin{array}{ll}\multicolumn{2}{l}{\textit{name-rel}}\\ \textsc{name} & \text{mary}\\ \textsc{named} & \boxed{2}\end{array}\right]\right\rangle \end{array}\right] \\ \textsc{ctxt} & \left[\begin{array}{ll} \textsc{spkr} & \boxed{3} \\ \textsc{addr} & \boxed{4}\end{array}\right]\end{array}\right]$$

The content can now be used directly as an interpretation for the utterance which can be passed to the dialogue move engine (DME).

2.2.4 Dialogue Management

Modifications to the DME were required in order to integrate the CMT classification used in the HPSG grammar with the dialogue system. At this early stage this was achieved by using a simplified update protocol based on that proposed in KOS(Ginzburg, 1996) and shown here:

```
LATEST-MOVE:  ask( Q )
       ->       push( Q, QUD ).

LATEST-MOVE:  assert( P ), max_qud( Q ),
              not( answers( P, Q ) ), not( in( P, COM ) )
       ->       push( P?, QUD ).

LATEST-MOVE:  assert( P ), max_qud( Q ),
              answers( P, Q )
       ->       pop( Q, QUD ), add( P, COM ),
                add( resolves( P, Q ), COM ).

...
```

This protocol alone gives no basis for actions by the system in response to any move by the user, effectively producing a monologue (rather than dialogue) system. The system in this state[1] can therefore be seen as an implementation of SHARDS within the TrindiKit framework, using the GoDiS DME to maintain a record of information state (IS).

2.2.5 Answerhood

SHARDS uses situation semantics to represent the content of utterances, and this was integrated into the dialogue system by passing the infons (encoded as complex Prolog entities) directly into the dialogue IS. This presented an interesting question of how to detect answerhood.

A simplistic solution was used whereby assertions were said to answer questions if their NUCLEUS values could be unified, thus ignoring any effects of quantification. Negation and truth were then regarded as quantifiers to allow, say, both `untrue(P)` and `true(P)` to be considered answers to `whether(P)` (see AVM [3]

[1] The system can be tested at `http://pc320.dcs.kcl.ac.uk:8080/tis`.

and AVM [4] below:

[3]
$$\begin{bmatrix} \text{PHON} & \langle\text{does,john,snore}\rangle \\ \text{CONT}\,|\,\text{MSG-ARG} & \begin{bmatrix} \text{QU-INDS} & \langle\ \rangle \\ \text{PROP} & \begin{bmatrix} \textit{proposition} & \\ \text{QUANTS} & \langle\ \rangle \\ \text{NUCLEUS} & \begin{bmatrix}\textit{snore-rel} & \\ \text{AGENT} & \boxed{1}\end{bmatrix} \\ \text{REST} & \left\{\begin{bmatrix}\textit{name-rel} & \\ \text{NAME} & \text{john} \\ \text{NAMED} & \boxed{1}\end{bmatrix}\right\}\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

[4]
$$\begin{bmatrix} \text{PHON} & \langle\text{no}\rangle \\ \text{CONT}\,|\,\text{MSG-ARG} & \begin{bmatrix} \textit{proposition} & \\ \text{QUANTS} & \left\langle\begin{bmatrix}\textit{untrue-rel} \\ \text{NUCL}\ \boxed{2}\end{bmatrix}\right\rangle \\ \text{NUCLEUS} & \boxed{2}\begin{bmatrix}\textit{snore-rel} & \\ \text{AGENT} & \boxed{1}\end{bmatrix} \\ \text{REST} & \left\{\begin{bmatrix}\textit{name-rel} & \\ \text{NAME} & \text{john} \\ \text{NAMED} & \boxed{1}\end{bmatrix}\right\}\end{bmatrix}\end{bmatrix}$$

This treatment may be found wanting as the other elements of the system become more realistic. One problem is the question of what restrictions to place on the REST values – a requirement for unification is too strong, but a weaker concept (such as consistency) may be difficult to implement.

2.2.6 Adding a Realistic Lexicon

SHARDS was originally designed and implemented as a "toy" system to demonstrate the capability of the ellipsis resolution techniques proposed in (Ginzburg et al., 2001a), and as such had no requirement for a realistic lexicon. Similarly GoDiS, due to its domain-specific nature, has a narrow domain-based lexicon.

In order for SHARDS to function as the input processor of an open-domain dialogue system, a realistic lexicon will certainly be required. The Oxford Advanced Learner's Dictionary of Current English (OALD – (Hornby, 1974)) is available in a machine-readable form (Mitton, 1992) and was used as an experiment into the feasibility of this. The OALD is a ~40,000 word dictionary of English distributed in a machine-readable format, which was converted into a suitable Prolog format by use of Perl scripts.

In addition to the addition of words themselves, some form of stemming or lemmatization is required for a dialogue system that is (in some way) to match questions with answers. It is vital to recognise that, say, *am* and *are* convey the same lemma *be*. The OALD contains morphological information which has been used to integrate stemming as part of the grammar.

The SHARDS grammar now functions with this lexicon with no discernible change in speed.

2.3 Work Proposed

2.3.1 Grammar Expansion

The limitations of the SHARDS system also include having a very limited grammar, which needs to be expanded to give more realistic coverage. Some work in this area is already being completed by Fernández Rovira (forthcoming); Dallas (forthcoming) and the results of this work will be incorporated when possible.

One problem that is likely to be encountered once the grammar becomes large is that of structural ambiguity. It may be that disambiguation methods such as the use of a stochastic grammar (see e.g. Brew, 1995) will be necessary (see chapter 5).

2.3.2 Synonyms

The use of a large realistic lexicon raises a problem with the need to recognise synonyms. In GoDiS, given the limited size of the lexicon, synonym relations are specified as lists:

```
synset( [ [flight], [flights], [plane], [fly], [airplane] ], plane ).
synset( [ [cheap], [second,class], [second] ], economy ).
synset( [ [first,class], [first], [expensive], [business,class] ], business )
```

With a large lexicon this is clearly going to become difficult. Some resources exist that specify links between related lexical concepts, and these are widely used in NLP applications to provide synonymity relations. One possibility might be WordNet (see Miller, 1995) – which has the advantage of being freely available – another CUP's CIDE+.

This approach might deal with examples such as the first line above where the related words are related in a domain-independent manner. The other two lines show that there may be a problem with domain-dependency: dictionaries are unlikely to regard, say, *cheap* and *second* as being synonymous (indeed, WordNet 1.6 gives no relation between them).

While this does not immediately impact the framework being developed, it will be an important requirement for any application so must be considered. It is hoped that a combination of dictionary-based relations with a small number of domain-specific, manually specified relations can be used.

2.3.3 Dialogue Management

The DME currently manages the dialogue IS – i.e. manages beliefs, questions under discussion and plans. The current implementation needs plans updating in order to be compatible with the HPSG/situation semantics approach.

Answerhood

It seems likely that the simple notion of answerhood described above will be insufficient for a realistic implementation. In particular, a distinction between answers that are *about* a question and those that *resolve* it, in the sense of (Ginzburg, 1996), may be needed. This is likely to be domain-dependent, and may also be dependent on the state of the current IS.

The notion of answerhood is important within the GoDiS approach, as the accommodation of unasked but planned questions onto the current shared IS is achieved by checking question-answer relevance. Whether the notion of a *relevant answer* as required by GoDiS coincides with a *resolving* or *about* answer remains to be seen.

Planning

Plans in GoDiS are currently lists of dialogue actions and are entirely domain-dependent. A first required step will be to convert the GoDiS plan schemata to be compatible with the HPSG/situation semantic approach now being used. The current plans are specified in a very simple and domain-dependent way – while this only becomes an issue for translating planned actions into output if a grammar-based output method is being used (see below), it is definitely an issue for updating information state on the basis of actions by the system (for example, the updating of QUD when the system asks a question).

Assuming a canned-text output method is being used, this requires one of two approaches: the specification of plans (or individual tasks within plans) in a situation-semantic way, together with a method of associating canned output strings with these complex tasks; or a simple task specification and output schema (similar to that already used), together with a method of associating these simple tasks with fully-formed semantic information state updates. Again, results from (Dallas, forthcoming) may prove useful here.

There will be overlap with the notion of answerhood: accommodation of plans into the IS is achieved by finding *relevant* unasked questions within possible plans.

More interesting planning issues will arise when clarification requests and underspecified representations are added to the system, as generic (domain-independent?) plans for utterance clarification and disambiguation will be required. These are discussed further in the relevant chapters below.

Ellipsis

Ellipsis resolution is performed by the SHARDS interface before reaching the DME. It seems possible that this resolution process would be better handled by the DME. It could then be potentially made dependent on IS contents, helping with possible disambiguation (see chapter 5). Although this approach is argued against by Lewin and Pulman (1995), the main grounds given are that ellipsis resolution must make use of syntactic and semantic representations not usually available within the IS – but this will not be the case in this implementation. However, they also point out that it makes the system less modular, making reconfiguration more difficult.

2.3.4 Output

Utterances made by the system must somehow be converted from intended conversational moves to surface text. While it might be possible to use the same grammar for generation as for interpretation, using e.g. a version of semantic head-driven generation, this is not a trivial exercise, and at least in the initial stages output will be produced using the approach already used by GoDiS (association of canned phrases with key semantic relations).

This in itself is not trivial and will involve some work to convert the GoDiS modules to be compatible with the AVM/situation semantic representation now used.

Many other changes to the implementation will be made during the course of this project. The main changes currently envisaged will result from the work on clarifications and underspecification. The next two chapters discuss these areas, together with the changes they might require.

Chapter 3

Clarifications

3.1 Background

3.1.1 The Importance of Clarifications

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

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

(2)
System: Would you like to travel via Paris or Amsterdam?
Caller: Paris?
System: (a) Yes, Paris.
(b) Paris, France.
(c) Paris is the quickest route, although Amsterdam is the cheapest.
(d) OK. Your ticket via Paris will be posted to you. Goodbye.

Any of responses (a)–(c), which correctly interpret the caller's move as a CR, might be regarded as useful to the caller: response (d), which incorrectly interprets it as an answer to the system's question, would not be acceptable under any circumstances. Which of (a)–(c) is preferred will depend on the reading intended. It is therefore important to investigate methods of identifying and correctly interpreting CRs.

This chapter describes investigation so far performed into the nature of CRs (form, reading and separation distance from the utterance being clarified), and then describes the questions that this investigation has raised and the work proposed in these question areas.

The rest of this section gives some background in proposed HPSG analysis of CRs.

3.1.2 HPSG Framework

Previous work by Ginzburg and Sag (2000) (hereafter G&S) and Ginzburg and Cooper (2001) (hereafter G&C) have examined some individual CR forms and given possible HPSG analyses for these forms.

G&S discuss *reprise interrogatives*, which they further classify into *echo* questions (those "resulting from mishearing a previous speech act" – see B's question in example (3)) and *reference* questions (those which "ask for clarification of the reference of some element in the immediately prior utterance" – see example (4)).

(3)
A: Did Jill phone?
B: Did JILL phone?

(4)
A: Did Jill phone?
B: Did WHO phone?

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

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

This analysis is then extended to two elliptical forms: *reprise sluices* and *elliptical literal reprises*. Sluices are elliptical wh-constructions (see Ross, 1969) – short wh-questions which receive a "sentential" interpretation, in this case an interpretation as a reprise question, as shown in example (5):

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

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

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

Resolution of these elliptical forms is achieved by allowing a CP to coerce a clarification question onto the list of questions under discussion (QUD) in the current dialogue context. This allows ellipsis resolution in the manner of of Ginzburg et al. (2001a) to give essentially the same reading as reprise questions.

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

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

3.1.3 Clarification-Source Separation

As described in chapter 1, an important question for research in this area is how long the surface structure of utterances is held in memory. The treatment of CRs outlined above requires such surface information about previous utterances in order to check phonological identity and query sign content.

One of the goals of the work described here is therefore to determine the maximum distance between a CR and the utterance being clarified (the *source* utterance) – *Clarification-Source Separation (CSS)* distance.

3.2 Work to Date

The aims of investigations undertaken so far have been to identify the available forms and readings for CRs, together with possible values for CSS distance, by corpus analysis. Initial analysis is complete and its results are described here.

Implementation of some forms and readings has also been carried out, resulting in a dialogue system capable of processing a limited range of CRs.

3.2.1 Corpus Analysis

As a first step towards a full theory of CR interpretation, corpus analysis was performed in order to gain information about which readings are available via which forms (with the aim of exhaustively categorising CR forms and readings) and about possible CSS distance.

A sub-portion of the British National Corpus (BNC) dialogue transcripts was used consisting of ∼150,000 words. To maintain a spread across dialogue domain, region, speaker age etc., this sub-portion was created by taking a 200-speaker-turn section from 59 transcripts.

All CRs within this sub-corpus were identified and tagged, using the markup scheme and decision process described in (Purver et al., 2001). Although initial identification of CRs was performed using SCoRE, the final search and markup were performed manually in order to ensure that all clarificational phenomena were captured.

The results of this work are given in detail in (Purver et al., 2001) and are presented briefly here.

3.2.2 SCoRE: a Corpus Search Tool

Initial corpus investigations into the distribution and form of CRs required a suitable corpus and search tool. The BNC, contains a ∼10 million word sub-corpus of English dialogue transcripts. Dialogues are classified by (amongst other categories) domain: about one fifth of the dialogue files are classified as *demographic* (non-context-governed) and the rest classified by context (e.g. *business, educational*).[1]

SARA, the search software supplied with the BNC, provides many functions including the ability to search for strings within the corpus, view results in context, and browse the corpus,[2] and a SARA server has been set up and tested at `bach.dcs.kcl.ac.uk`. However, SARA does not fulfil all the requirements necessary for searching for and browsing CRs. A list of requirements was defined as follows:

[1] See (Burnard, 2000) for a full description of the classification system.

[2] A full specification for SARA is given in (Dodd, 1997).

- Ability to search for word strings (including punctuation)
- Ability to search for repeated strings of user-specifiable length and separation
- Ability to search on a part-of-speech (PoS) basis
- Ability to search on the basis of sentences or speaker turns
- Ability to restrict searches/results to files of particular types (e.g. written/spoken)
- Ability to view matches in context of user-specifiable depth
- Ability to browse dialogue files via a user-friendly interface

SARA provides only the first and fifth of these requirements (PoS searches are available, but only as refinements of prior word searches), so a dedicated tool, SCoRE[3], was written. A full description of and manual for SCoRE is available in (Purver, 2001).

3.2.3 Clarification Forms

The following forms have been identified as possible means for CRs. While this list cannot be guaranteed exhaustive, a markup scheme based on these forms has been shown to cover the CRs encountered in a corpus of dialogue, as detailed in section 3.2.5 below. This section lists the forms identified, and illustrates them with examples. All examples have been taken from the BNC.

Non-Reprise Clarifications

Unsurprisingly, speakers have recourse to a non-reprise[4] form of clarification. In this form, the nature of the information being requested by the CR initiator is spelt out for the addressee. Utterances of this type thus often contain phrases such as *"do you mean. . . "*, *"did you say. . . "*, as can be seen in examples (7) and

[3] Originally, "**S**earch a **Co**rpus for **R**egular **E**xpressions".

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

(8).

(7)[5]
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.

(8)[6]
Leon: Erm, your orgy is a food orgy.
Unknown: **What did you say?**
Leon: Your type of orgy is a food orgy.

Reprise Sentences

Speakers can form a CR by echoing or repeating[7] a previous utterance in full, as shown in example (9). This form corresponds to G&S's *reprise interrogative*.

(9)[8]
Orgady: I spoke to him on Wednesday, I phoned him.
Obina: **You phoned him?**
Orgady: Phoned him.

This form appears to be divisible into two sub-categories, *literal* (as in example (9) above) and *wh-substituted* reprise sentences, as illustrated by example (10).

(10)[9]
Unknown: He's anal retentive, that's what it is.
Kath: **He's what?**
Unknown: Anal retentive.

Reprise Sluices

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

(11)[10]
Sarah: Leon, Leon, sorry she's taken.
Leon: **Who?**
Sarah: Cath Long, she's spoken for.

There may be a continuum of forms between *wh-substituted reprise sentences*

[5] BNC file KP4, sentences 521–524

[6] BNC file KPL, sentences 524–526

[7] Repeats need not be verbatim, due to the possible presence of phenomena such as anaphora and VP ellipsis, as well as changes in indexicals as shown in example (9).

[8] BNC file KPW, sentences 463–465

[9] BNC file KPH, sentences 412–414

[10] BNC file KPL, sentences 347–349

and *reprise sluices*. Consider the following exchange (12):

(12)[11]	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.[12]

Reprise Fragments

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

	Lara:	There's only two people in the class.
(13)[13]	Matthew:	**Two people?**
	Unknown:	For cookery, yeah.

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

	Ben:	No, ever, everything we say she laughs at.
(14)[14]	Frances:	**Who Emma?**
	Ben:	Oh yeah.

As these examples appeared to be interchangeable with the plain fragment alternative (in example (14), *"Emma?"*), they were not distinguished from fragments in our classification scheme.

Gaps

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

[11] BNC file KSV, sentences 363–364

[12] While the current exercise has not highlighted it as an issue, we note that a similar continuum might be present between literal reprises and reprise fragments. One approach in the face of this indeterminacy might be to conflate these forms – further analysis of the results given in this paper may indicate whether this is desirable.

[13] BNC file KPP, sentences 352–354

[14] BNC file KSW, sentences 698–700

this component – see example (15).

(15)[15]
Laura: Can I have some toast please?
Jan: **Some?**
Laura: Toast

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

Gap Fillers

The *filler* form is used by a speaker to fill a gap left by a previous incomplete utterance. Its use therefore appears to be restricted to such contexts, either because a previous speaker has left an utterance "hanging" (as in example (16)) or because the CR initiator interrupts.

(16)[16]
Sandy: if, if you try and do enchiladas or
Katriane: Mhm.
Sandy: erm
Katriane: **Tacos?**
Sandy: tacos.

Conventional

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

(17)[17]
Anon 2: Gone to the cinema tonight or summat.
Kitty: **Eh?**
Anon 2: Gone to the cinema

3.2.4 Clarification Readings

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

[15] BNC file KD7, sentences 392–394
[16] BNC file KPJ, sentences 555–559
[17] BNC file KPK, sentences 580–582

Clausal

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

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

An attribute-value matrix (AVM) skeleton for the semantic content of an HPSG sign corresponding to this reading (according to G&C's analysis) is shown below as AVM [5]. It represents a question[18], the propositional content of which is the conversational move made by the source utterance (shown here as being of type *illoc(utionary)-rel(ation)* – possible subtypes include *assert, ask*) together with the message associated with that move (e.g. the proposition being asserted). The parameter set being queried can be either a constituent of that message (as would be the case in a sluice or wh-substituted form, where the CR question is the wh-question *"For which X are you asserting ..."*) or empty (as would be the case in a fragment or literal reprise form, where the CR question is the polar question *"Are you asserting ..."*).

$$[5] \quad \left[\begin{array}{ll} \textit{question} & \\ \text{PARAMS} & \{\boxed{1}\} \text{ or } \{\ \} \\ \text{PROP} \mid \text{SOA} & \left[\begin{array}{ll} \textit{illoc-rel} & \\ \text{MSG-ARG} & \left[\ldots\boxed{1}\ldots\right] \end{array}\right] \end{array}\right]$$

Constituent

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

This reading corresponds roughly to *"What/who is X?"* or *"What/who do you mean by X?"*, as shown in AVM [6], a description of the content that would be given by G&C's analysis. This shows a question whose propositional content is the relation between a sign (a constituent of the source utterance) and its

[18] We adopt here the version of HPSG developed in G&S, wherein questions are represented as semantic objects comprising a set of parameters (empty for a polar question) and a proposition. This is the feature-structure counterpart of a λ-abstract wherein the parameters are abstracted over the proposition.

semantic content. The abstracted parameter is the content.[19]

$$[6]\quad \begin{bmatrix} \textit{question} & \\ \text{PARAMS} & \{\boxed{2}\} \\ \text{PROP}\mid\text{SOA} & \begin{bmatrix}\textit{content-rel} & \\ \text{SIGN} & \boxed{1} \\ \text{CONT} & \boxed{2}\end{bmatrix}\end{bmatrix}$$

Lexical

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

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

Corrections

The correction reading appears be along the lines of *"Did you intend to utter X (instead of Y)?"*. No full analysis for this reading has yet been proposed.

3.2.5 Results

Results are given in in detail in (Purver et al., 2001) and a brief summary is given here.

- CRs were found to make up between 3 and 4% of sentences. This is a significant proportion, giving support to the claim that processing of CRs is important for a dialogue system.

- The coverage of the corpus by the forms and readings listed above is good, with only 0.5% of CR readings (2 sentences) and about 1.5% of CR forms (6 sentences) being classified as *other*.[20]

[19] It seems clear that the value of the SIGN attribute must be underspecified in some way – at least in its CONTENT value.

[20] The readings not covered were all expressing surprise, amusement or outrage at a previous utterance (rather than requesting clarification directly), and were all of the reprise fragment or conventional form. Our intuition is that these readings can be treated as clausal readings with a further level of illocutionary force given by use in context. Of the 2 sentences left unclassified for form, one appears to be an unusual conventional reading, and one an interesting example of a literal reprise of an unuttered but implied sentence.

- Of the non-conventional reprise forms, only the reprise fragment appears to *require* an analysis that gives a constituent reading.
- The gap and filler forms appear only to be used with a lexical reading, although few examples of these forms were encountered.
- The maximum CSS distance observed was 15 sentences. Only one example of this distance was observed, and one example of distance 13 – otherwise all CSS distances were below 10 sentences.[21] The vast majority of CRs had a CSS distance of one (i.e. were clarifying the immediately preceding sentence – see figure 3.1), and over 96% had a distance of 4 or less.

	`non`	`lit`	`sub`	`slu`	`frg`	`gap`	`fil`	`wot`	`oth`	Total
`cla`	4.3	4.8	1.0	10.7	25.2	0	0	0	0.5	46.5
`con`	7.6	0	0	0	1.7	0	0	5.3	0	14.5
`lex`	0.7	0	2.6	2.1	0.2	0.5	3.8	25.0	0	35.0
`cor`	1.0	0.5	0	0	1.0	0	0	0	0	2.4
`oth`	0	0	0	0	1.0	0	0	0.5	0	1.4
Total	13.6	5.3	3.6	12.8	29.1	0.5	3.8	30.7	0.5	100.0

Table 3.1: CR form and type as percentage of CRs – all domains

The (tentative) conclusions drawn were:

- An automated dialogue system which can deal with fragments, sluices and reprise sentences, together with conventional and non-reprise CRs, could give reasonable coverage of expected dialogue. Fillers and especially gaps make up only a small proportion.
- The high proportion of lexical readings suggests that a detailed analysis of this phenomenon will be required.
- An utterance record with length of the order of ten sentences would be sufficient to allow a dialogue system to process the vast majority of CRs.
- Many readings are available for some forms. This implies that disambiguation between readings will be important for a dialogue system.

[21] The sentences with CSS distances of 13 and 15 were anomalous in that they occurred in a classroom context where many speakers were taking turns, increasing the apparent distance.

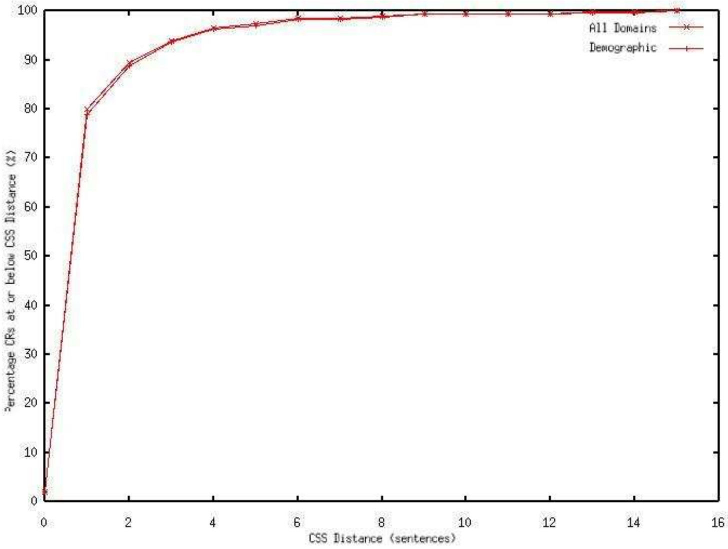

Figure 3.1: Percentage of CRs vs. CR-Source Separation Distance

3.2.6 Implementation

As a next step the treatment of CE as proposed by Ginzburg and Cooper (2001) was implemented within the SHARDS framework. This allowed clausal and constituent readings to be produced from bare fragment inputs by the user.[22]

An example IS is shown here, after a clausal clarification move (but before updating of QUD). The source utterance being clarified was assertion *"John snores."* (so the QUD left from the previous state can be seen to the the question of *whether John snores*). The CR utterance was *"John?"*, which has been

[22] The system can be tested at `http://pc320.dcs.kcl.ac.uk:8080/ce`.

interpreted as asking the question *did you assert that John snores?*

[7]
$$\begin{bmatrix} \text{COM} & \{\ \} \\ \text{QUD} & \left\langle \begin{bmatrix} \textit{question} & \\ \text{QU-INDS} & \{\ \} \\ \text{PROP} & \boxed{2} \begin{bmatrix} \text{QUANTS} & \langle\ \rangle \\ \text{NUCLEUS} & \begin{bmatrix} \textit{snore-rel} & \\ \text{AGENT} & \boxed{1} \end{bmatrix} \\ \text{REST} & \begin{bmatrix} \textit{name-rel} & \\ \text{NAME} & \text{john} \\ \text{NAMED} & \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \\ \text{LU} & \begin{bmatrix} \text{SPEAKER} & \boxed{3} \\ \text{MOVES} & \left\{ \begin{bmatrix} \textit{ask} & \\ \text{UTT} & \boxed{3}\ \text{usr} \\ \text{ADD} & \text{sys} \\ \text{MSG-ARG} & \begin{bmatrix} \text{QU-INDS} & \{\ \} \\ \text{PROP}\,|\,\text{NUCLEUS} & \begin{bmatrix} \textit{assert-rel} & \\ \text{UTT} & \text{sys} \\ \text{ADD} & \text{usr} \\ \text{MSG-ARG} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix} \\ \text{NIM} & \{\ \} \end{bmatrix}$$

This analysis of CE allows any fragment with a semantic index to function as a CE utterance, and this leads to CE readings being attributed in too many cases. To reduce the scale of the problem, the order of processing within SHARDS was arranged such that all other interpretations of utterances were attempted first, and MAX-QUD coercion (leading to CE resolution) only attempted if other interpretations failed. A more sophisticated approach might result from research into disambiguation of possible CRs (see chapter 5).

A major limitation of the current implementation is that only an immediately preceding utterance can be clarified, as no memory for surface forms is implemented.

3.3 Work Proposed

The immediate aim of further research is to confirm and extend the corpus work (partly by experiments with human subjects), to provide HPSG analyses for the forms and readings identified, and to extend the implementation to cover the new forms and readings.

3.3.1 Corpus Work

The corpus markup process has been performed by only one (expert) user – results must now be confirmed by comparing with those obtained by naive users, using e.g. the kappa statistic (Carletta, 1996) to assess reliability.

3.3.2 Experiments

Experiments using human subjects will be used to confirm CSS distance results. The form of the experiments is not yet determined, but might involve the use of a confederate to guide dialogue and introduce CRs of varying forms at varying CSS distances.

3.3.3 CR Readings

HPSG analyses are required for the lexical and correction readings.

Lexical

The lexical reading must refer in some way to the phonological form of the source utterance. A tentative proposal for the semantic content is as shown in AVM [8]. Here, the content shown is a question whose propositional content is the proposition that a speaker uttered a word with a particular phonological form – this is the parameter being abstracted to form the question.

$$[8]\quad \begin{bmatrix} \textit{question} & \\ \text{PARAMS} & \{\boxed{1}\} \text{ or } \{\ \} \\ \text{PROP}\,|\,\text{SOA} & \begin{bmatrix} \textit{utter-rel} & \\ \text{SIGN} & \begin{bmatrix}\text{PHON} & \boxed{1}\end{bmatrix}\end{bmatrix} \\ \text{CTXT} & \begin{bmatrix}\text{PREV-UTT}\,|\,\text{PHON} & \langle \ldots \boxed{1} \ldots \rangle\end{bmatrix}\end{bmatrix}$$

Correction

More work is required before an analysis can be proposed for this reading. It may be that corrections can in fact have clausal, constituent or lexical sub-type, so it must be taken into account that this may in fact not be a separate reading but a particular usage of those already established.

3.3.4 CR Forms

HPSG analyses are required for the gap, filler and conventional forms. Before these analyses can be completed, further work must establish the readings that can be taken by these forms.

Gaps

The gap form poses an interesting puzzle as it must refer to a part of the surface form of the source utterance immediately after that used to form the CR. If only lexical readings are required, a tentative proposal might be as follows:

$$[9]\quad \begin{bmatrix} \textsc{phon} & \boxed{1} \\ \textsc{cont} & \begin{bmatrix} \textit{question} & \\ \textsc{params} & \{\boxed{2}\} \\ \textsc{prop}\mid\textsc{soa} & \begin{bmatrix} \textit{utter-rel} & \\ \textsc{sign} & \begin{bmatrix} \textsc{phon} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \textsc{ctxt} & \begin{bmatrix} \textsc{prev-utt}\mid\textsc{phon} & \langle \ldots \boxed{1}, \boxed{2} \ldots \rangle \end{bmatrix} \end{bmatrix}$$

Gap Fillers

The analysis of the filler form is problematic and more thought is required before any proposal is made. The question posed appears to be of the form (for a lexical reading) *"Did you intend to utter X next?"*.

Conventional

Analysis of the conventional form is not trivial as it must be compatible with several readings (a possible analysis for a lexical reading might be as shown in AVM [10]).

It seems desirable for the representation to be stored in the lexicon, as the form can only be used with a finite number of conventionalised words (this fits with the approach to conventionalised conversational moves such as greetings outlined in (Ginzburg et al., 2001b)). However, it does not seem sensible to introduce separate lexical entries for different readings produced by the same word.

$$[10]\quad \begin{bmatrix} \textsc{phon} & \langle \text{eh?} \rangle \\ \textsc{cont} & \begin{bmatrix} \textit{question} & \\ \textsc{params} & \{\boxed{1}\} \\ \textsc{prop}\mid\textsc{soa} & \begin{bmatrix} \textit{utter-rel} & \\ \textsc{sign}\mid\textsc{phon} & \boxed{1} \end{bmatrix} \end{bmatrix} \\ \textsc{ctxt} & \begin{bmatrix} \textsc{prev-utt}\mid\textsc{phon} & \boxed{1}\langle \ldots \rangle \end{bmatrix} \end{bmatrix}$$

3.3.5 Implementation

Grammar

SHARDS must be extended to include any new CR forms and readings, so that the grammar can produce reasonable interpretations of all CR moves.

Information State

Some form of memory for surface structure must be introduced into the dialogue IS in order to allow prior utterances to be the subject of clarification. The most straightforward approach to this would be the introduction of a PREVIOUS-UTTERANCE stack or set (of limited size) which could act as a record of full utterance signs, including phonology and syntax as well as semantic content. This might be achieved by an extension of the existing LATEST-MOVE.

Dialogue Management

Once this is in place, strategies for producing and for dealing with CRs must be designed and implemented.

It is hoped that such strategies could be incorporated as domain-independent plans which could be executed by the DME as *nested* sub-plans (within the main domain-specific task plan). Although any clarification strategy is likely to need to draw on domain knowledge (in order to supply the required clarifying information), the strategy itself may be a general one and could follow the Utterance Processing Protocol (UPP) outlined in G&C:

```
if ( fully_grounded_interpretation( U, S ) )
  ->
    % treat as normal utterance
    latest_move = S
    (proceed as usual)
else
  ->
    % attempt to interpret as clarification question
    coerce_max_qud( U )
    repeat
else
  ->
    % produce clarification question to aid interpretation
    clarify( U )
```

Any strategy must deal with questions currently under discussion, allowing the system to return to them once the clarification has been successfully performed and utterance content updated.

It is also clear that such a strategy will require some degree of underspecification of the HPSG utterance representations. This is discussed in the next chapter.

Chapter 4

Underspecification

4.1 Background

It is often the case that utterances cannot immediately be fully interpreted. This may be due to, for example, the presence of unknown words, reference to unknown or insufficiently specified entities, or ambiguity caused by multiple possible parse structures or ways of resolving ellipsis.

In many of these cases it seems likely that the human processing system does produce some partial representation of the utterance, which must be underspecified in some way corresponding to the unknown or ambiguous quantity. Work in psycholinguistics suggests that such underspecified representations are common (see e.g. Poesio, 1996). In order to produce CRs such as that in example (14) above (repeated here as example (18)), and then successfully integrate the response, it seems that the CR initiator must have processed the source sentence in some way sufficient to produce a representation of its meaning in which (only) the referent of *she* is unknown. This representation can then be updated to include the correct referent *Emma* when it is confirmed.

(18)[1]
Ben: No, ever, everything we say she laughs at.
Frances: **Who Emma?**
Ben: Oh yeah.

Computational approaches to dialogue have not generally attempted to construct such underspecified representations, relying instead on robust processing techniques to extract fully specified elements (although these may well correspond only to parts of the utterance being processed). However, if a genuinely linguistic processing method is to be used, an ability to build underspecified representations is highly desirable – the alternative is to reject any utterances that cannot be fully interpreted.

Similar techniques have been developed for related fields. In speech recognition, the presence of multiple possible solutions is the rule, rather than the

[1] BNC file KSW, sentences 698–700

exception. Recognizers often give a *N-best* output, returning the N most likely possible word strings corresponding to a sound signal. A more compact way of representing these possible strings is by use of a *lattice* – a network or graph with edges corresponding to words with associated probabilities.

Probabilistic parsers usually use the N-best approach, and little work appears to have been done on more compact or efficient representations. It seems possible that a lattice-like approach could be used, perhaps taking advantage of the already compact chart representation used during parsing.

Milward (1999, 2000) proposes a chart-based approach to semantics in which the output of a parser becomes a compact representation of partial or multiple possible interpretations of an utterance. This approach is based around *dependency grammar* (see Milward, 1994) – however, it is not obvious how to translate it to an HPSG approach, and in particular to extend it to include levels other than semantics.

Underspecified semantics has received a great deal of attention, but mostly as an attempt to deal with quantifier scope ambiguity (although some attention has also been given to anaphora) rather than the issues described here. Richter and Sailer (1999) have shown that it is possible to implement one of these approaches in HPSG: they give a treatment for *hole semantics* (see Bos, 1995) – a semantic underspecification language designed to represent quantifier scope and PP attachment ambiguities. It might be possible to apply a similar approach to some of the types of ambiguity discussed here. Reyle (1996) also presents a formal theory of underspecification in DRT which might be applicable to dialogue information states.

4.2 Work to Date

Work to date has consisted of identifying the following areas which need to be treated, together with some possible approaches that might be used. A preliminary implementation of one approach has been started.

4.2.1 Unknown Referents

One interesting question is how a partially grounded utterance is represented in a CP's memory. The phenomenon of reprise sentences suggests that such a partial representation does exist: in example (19) below, the (unnamed) CR initiator has at least managed to understand the content of the source utterance to have a form *"Don't talk about X"* – although he has not managed to ground the component *it* in context and thus does not know its content (reference).

(19)[2]
Danny: Just shove off don't talk about it alright.
Unknown: **Talk about what?**
Danny?
Don't talk about what?

A similar phenomenon can be observed in example (20), where Cassie has interpreted Peter's source utterance but cannot ground a particular aspect of it (its temporal aspect) in the current context without clarifying it:

(20)[3]
Peter: `<unclear>` go back to school.
Cassie: **What now?**
Peter: Yeah.

It is not entirely clear what representation such a partially grounded utterance might be given. A possibility is to leave semantic indices uninstantiated, allowing them to be unified with known (or new) referents at a later stage (i.e. after clarification). This would license representations along the lines of AVM [11], where the time parameter associated with the required outcome (*that you go back to school*) has not been fully instantiated (and thus is not re-entrant with a member of the C-INDICES set. It is shown here as an uninstantiated Prolog-style variable X.[4] The contrasting fully instantiated equivalent is shown as AVM [12]: here the time parameter has been instantiated and is re-entrant with the time of

[2] BNC file KPA, sentences 950–953

[3] BNC file KP4, sentences 378–380

[4] This form of uninstantiation may not fit with current formal HPSG theory, so an alternative must be investigated.

the utterance (*now*).

$$[11]\quad \left[\begin{array}{ll} \text{CONT} & \left[\begin{array}{ll} \multicolumn{2}{l}{\textit{outcome}} \\ \text{SOA} & \left[\begin{array}{ll} \text{T-PARAM} & \left[\begin{array}{ll}\text{INDEX} & \boxed{1}\text{X}\end{array}\right] \\ \text{NUCL} & \left[\begin{array}{ll} \multicolumn{2}{l}{\textit{go-back-to-school-rel}} \\ \text{AGENT} & \boxed{2} \\ \text{TIME} & \boxed{1} \end{array}\right] \end{array}\right] \end{array}\right] \\ \text{CTXT} & \left[\begin{array}{ll} \text{C-INDICES} & \{\boxed{2}\} \\ \text{ADDR} & \boxed{2} \end{array}\right] \end{array}\right]$$

$$[12]\quad \left[\begin{array}{ll} \text{CONT} & \left[\begin{array}{ll} \multicolumn{2}{l}{\textit{outcome}} \\ \text{SOA} & \left[\begin{array}{ll} \text{T-PARAM} & \left[\begin{array}{ll}\text{INDEX} & \boxed{1}\end{array}\right] \\ \text{NUCL} & \left[\begin{array}{ll} \multicolumn{2}{l}{\textit{go-back-to-school-rel}} \\ \text{AGENT} & \boxed{2} \\ \text{TIME} & \boxed{1} \end{array}\right] \end{array}\right] \end{array}\right] \\ \text{CTXT} & \left[\begin{array}{ll} \text{C-INDICES} & \{\boxed{1},\boxed{2}\} \\ \text{UTT-TIME} & \boxed{1} \\ \text{ADDR} & \boxed{2} \end{array}\right] \end{array}\right]$$

This approach might provide a reasonable basis for the production of CRs (any uninstantiated variables can form the abstracted parameters of a clarification question). However, it raises the question of what level of uninstantiation is acceptable within a sign and how this level can be determined if it is (as seems inevitable) context-dependent.

4.2.2 Unknown Words

The problem posed by out-of-vocabulary (OOV) words is a long-standing problem in speech recognition technology (see e.g. Hetherington, 1994). However large the lexicon is made, new words will always be encountered, and this causes problems for speech recognisers as they will have no corresponding entries in either the acoustic or language models.

Gallwitz et al. (1996) propose a category-based method (developed further in (Boros et al., 1997)) in which categories for OOV words are estimated based upon probabilities determined from corpus frequency data. It seems possible that a similar approach could be followed in parsing, substituting a generic word entry of a suitable part-of-speech category (determined either from word surface form or from surrounding context – i.e. the requirement to produce a complete parse for the utterance).

An approach of this type would necessarily leave the lexical semantics of the

OOV word underspecified in some way, as shown in AVM [13] (for an OOV verb) and AVM [14] (for an OOV noun). It might be possible to represent this formally using hole semantics.

[13]

$$\begin{bmatrix} \textsc{phon} & \langle \text{john,gazumped,mary} \rangle \\ \textsc{cont} & \begin{bmatrix} \textit{proposition} & \\ \textsc{soa} & \begin{bmatrix} \textsc{nucl} & \begin{bmatrix} \textit{dyadic-rel} & \\ \textsc{reln} & \text{X} \\ \textsc{agent} & \boxed{1} \\ \textsc{object} & \boxed{2} \end{bmatrix} \\ \textsc{rest} & \left\langle \begin{bmatrix} \textit{name-rel} & \\ \textsc{name} & \text{john} \\ \textsc{named} & \boxed{1} \end{bmatrix}, \begin{bmatrix} \textit{name-rel} & \\ \textsc{name} & \text{mary} \\ \textsc{named} & \boxed{2} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

[14]

$$\begin{bmatrix} \textsc{phon} & \langle \text{john,likes,kumquats} \rangle \\ \textsc{cont} & \begin{bmatrix} \textit{proposition} & \\ \textsc{soa} & \begin{bmatrix} \textsc{nucl} & \begin{bmatrix} \textit{dyadic-rel} & \\ \textsc{reln} & \text{like} \\ \textsc{agent} & \boxed{1} \\ \textsc{object} & \boxed{2} \end{bmatrix} \\ \textsc{rest} & \left\langle \begin{bmatrix} \textit{name-rel} & \\ \textsc{name} & \text{john} \\ \textsc{named} & \boxed{1} \end{bmatrix}, \begin{bmatrix} \textit{rest-rel} & \\ \textsc{index} & \boxed{2} \\ \textsc{restr} & \text{Y} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

This approach seems to fit well with work in psycholinguistics and cognitive science (see e.g. Poesio, 1996). It also provides a basis from which to produce CRs, as the uninstantiated elements of the content may be queried.

4.2.3 Implementation

This category-based approach has been implemented within the dialogue system in a preliminary fashion and shown to give successful interpretations for sentences such as those shown in AVM [13] and AVM [14] above.

Problems may well be encountered as utterances become longer, as the grammar used is expanded, and as more OOV words are encountered per utterance. Further testing will reveal whether this is the case – if so, the parser must be limited (by time, say, or number of OOV words) or must use some disambiguation strategy.

4.2.4 Ambiguity

A further potential source of underspecification is the presence of ambiguity. Possible sources of ambiguity include lexical ambiguity (multiple possible lexical entries for the same surface form), structural ambiguity (multiple possible parses), scope ambiguity (multiple possible scopings for semantic quantifiers) and contextual ambiguity (multiple possible ellipsis resolution derivations). In each case the ambiguity leads to more than one possible (fully grounded) interpretation for a given utterance.

An example of contextual ambiguity is shown here in example (21) – B's response could be interpreted as a tentative short answer (*"could it be Mary that likes Mary best?"*) or a reprise question (amongst other readings, *"Who is Mary?"*).

(21)
A: Who do you suppose likes Mary best?
B: **Mary?**

The simplest method is to represent utterances as a set or disjunction of all possible fully specified signs – this is essentially the approach used by most parsers and is termed a *parse forest*. In the case of probabilistic parsing, the set becomes an ordered N-best list.

This has the disadvantage of producing a inefficient representation which gives no information about any commonality between the elements of the set (or members of the disjunction). Representing this commonality seems important – in many cases, ambiguities might give rise to differences in semantics which should not affect the response of the system, perhaps because the differences do not affect the core part of the message. It might be possible to reintroduce this commonality via logical operations on the set of signs, if the signs were represented in a logical language such as RSRL (Richter, 2000). A unification-based approach might also be viable.

A more efficient method might be one based around the chart approach of Milward (2000) (but it must be adapted to HPSG and extended to include levels other than semantics), the packed approach used in Verbmobil (see Worm and Rupp, 1998), or the hole semantics approach of Richter and Sailer (1999).

Another approach might be to use less specific representations or *sign descriptions* rather than maximally specific signs. HPSG grammar rules are specified in this way – the hierarchical type system allows rules and constraints to be specified at more or less specific levels. Whether such a system of descriptions suitable for representation of ambiguity could be compatible with one suitable for the existing syntactic requirements of the grammar and parser remains to be seen.

4.3 Work Proposed

The next stage is to investigate the possible approaches mentioned above thoroughly and assess their suitability by implementation and testing.

4.3.1 Implementation

Grammar and IS

The grammar and IS will require substantial revision to produce the intended underspecified representations.

It is not clear how to implement a category-based approach in HPSG, especially in such a manner as to prevent explosion in the number of possible parses produced. This problem may become acute when several sources of underspecification are present in a single utterance. Some deterministic methods may be available (unknown words, for instance, must be limited to the open PoS categories such as noun, verb etc.), although probabilistic methods will also be investigated. Due to limited resources, any testing probabilistic methods may be limited to testing of frameworks with pre-determined probability values, rather than full training from corpus data.

Dialogue Management

A strategy for grounding and subsequent processing of underspecified utterances must be designed and implemented. This will have significant interplay with any clarification strategy, as decisions must be taken as to whether clarifications are required depending on the level of underspecification.

A grounding strategy must instantiate utterances as fully as possible, anchoring referents to contextual objects, but must be robust enough to allow underspecification when necessary. A subsequent processing strategy must include the ability to clarify utterances which are not sufficiently instantiated, but also to incorporate useful utterances (and possibly useful elements of under-instantiated utterances) into the IS.

4.3.2 Possible Problems

It is not clear how to integrate the approaches outlined above with the constituent and/or clausal CR readings outlined in chapter 3 above. Firstly, some link between content and original word/sign is required. In order to allow a CR which queries the content of a sign and to subsequently integrate the clarificational response to produce a fully instantiated interpretation, some record of the relation between surface form and the uninstantiated elements of the content must be maintained.

One approach might be to use a list of *non-integrated signs* similar to the *non-integrated move* list used in the dialogue information state approach of Larsson et al. (2000), producing structures of the form shown as AVM [15] below. Whether such a list would be more suitably represented as part of utterance context (as shown below), as an alternative DTRS list, or as part of dialogue information state remains to be seen.

$$\text{[15]}\quad \begin{bmatrix} \text{PHON} & \langle \text{john,likes,jezebel} \rangle \\ \text{CONT} & \begin{bmatrix} \textit{proposition} & \\ \text{SOA} & \begin{bmatrix} \text{NUCL} & \begin{bmatrix} \textit{dyadic-rel} & \\ \text{RELN} & \text{like} \\ \text{AGENT} & \boxed{1} \\ \text{OBJECT} & \boxed{2} \end{bmatrix} \\ \text{REST} & \left\langle \begin{bmatrix} \textit{name-rel} & \\ \text{NAME} & \text{john} \\ \text{NAMED} & \boxed{1} \end{bmatrix}, \boxed{3} \begin{bmatrix} \textit{name-rel} & \\ \text{NAME} & \text{jezebel} \\ \text{NAMED} & \boxed{2}\text{X} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \\ \text{CTXT} & \begin{bmatrix} \text{C-INDICES} & \{\boxed{1}\} \\ \text{NIS} & \left\{ \begin{bmatrix} \textit{sign} & \\ \text{PHON} & \langle \text{jezebel} \rangle \\ \text{CAT} & \begin{bmatrix} \textit{name} \end{bmatrix} \\ \text{CONT} & \boxed{3} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix}$$

It may be that this record could be combined with the functions of the PREVIOUS-UTTERANCE feature mooted in the previous chapter to provide an integrated theory.

Chapter 5

Disambiguation

5.1 Background

5.1.1 The Need to Disambiguate

Much of the completed and proposed work described above indicates the possible requirement of disambiguation techniques.

Firstly, as noted in chapter 2, the extension of the SHARDS grammar is likely to result in an explosion in the number of possible solutions given by the parser. The number of possible parses typically produced by a grammar powerful enough to give reasonable coverage can be extremely high for some sentences – (Martin et al., 1987) report 455 parses for:
"List the sales of the products produced in 1973 with the products produced in 1972."

Secondly, the treatment outlined in chapter 3 allows for multiple possible readings of CRs (e.g. clausal, constituent and lexical readings for a reprise fragment). Indeed, many CR utterances might be ambiguous between forms – a single-word utterance might be a reprise fragment or a gap CR. In addition, even CRs with a known reading might be ambiguous as to which part of the source utterance is being clarified (consider a reprise sluice *"Who?"* clarifying an utterance mentioning more than one human).

Thirdly, elliptical expressions can be ambiguous between type of ellipsis, as noted in chapter 4 – a single-word utterance might be interpretable as a short answer or as a reprise fragment CR.

While it may be desirable to represent truly ambiguous utterances as in some way underspecified (see chapter 4 above), it seems more useful to attempt to disambiguate cases which do not appear ambiguous to humans.

No disambiguation work has so far been attempted as part of this project.

5.2 Work Proposed

5.2.1 Stochastic Grammars

The traditional approach to parse disambiguation (in order to rule out spurious structural ambiguity) has been the use of stochastic grammars, in which grammar rules are associated with probabilities. These probabilities are summed as rules are applied, to produce the total probability of a parse tree. The probabilities may be conditioned on other rules and/or on the words themselves.

Given the extensive nature of previous research in this topic, it is not proposed to perform further research in this area, but some work may be required to produce a suitable stochastic grammar for the implementation.

Brew (1995) has proposed a stochastic version of HPSG that might be used. As training is time-consuming, a version using hand-coded probability values might be sufficient.

5.2.2 Clarification Form

The gap CR of example example (15), repeated here as example (22), is a single-word utterance *"Some?"*. According to the current proposed analysis of CRs, it could also be interpreted as a clausal, constituent or lexical CR querying the meaning or form of the word *some*.

(22)[1]
Laura: Can I have some toast please?
Jan: **Some?**
Laura: Toast

However, it seems that intonation is used to signal the gap form, allowing a CP to uniquely determine that this form is being used (and that the next word *toast* is actually the subject of the CR). If this is the case, it should be possible to use intonation to disambiguate CR form within the grammar. This would of course require intonation to be made a part of the HPSG sign representation.

While some work has been done on the use of intonation to assign conversational moves (see e.g. Kowtko et al., 1991) and to constrain speech recognition (see e.g. Taylor et al., 1997), there has been little within formal theories of grammar. HPSG's multi-level nature would seem to lend itself naturally to this kind of addition: intonational features could be used as further constraints on phrase types.

Examples 16 and 17 show the kind of approach envisaged. Signs of type *spoken* might have a PHON attribute which is a list of *spoken-word* AVMs, containing both word form and intonational information (represented here in autosegmental notation). Constraints on phrase type might then be specified in terms of

[1] BNC file KD7, sentences 392–394

intonational information (e.g. the boundary tone of the last word on the PHON list).

$$\text{[16]}\quad \begin{bmatrix} \textit{spoken \& gap-cr-form} & \\ \textsc{phon} & \left\langle \begin{bmatrix} \textit{spoken-word} & \\ \textsc{lex} & \text{some} \\ \textsc{inton} & \langle \textsc{h*,h,l\%} \rangle \end{bmatrix} \right\rangle \end{bmatrix}$$

$$\text{[17]}\quad \begin{bmatrix} \textit{spoken \& frag-cr-form} & \\ \textsc{phon} & \left\langle \begin{bmatrix} \textit{spoken-word} & \\ \textsc{lex} & \text{some} \\ \textsc{inton} & \langle \textsc{l+h*,l,h\%} \rangle \end{bmatrix} \right\rangle \end{bmatrix}$$

If, as seems possible, intonational constraints are not "hard" constraints, a probabilistic approach might provide a suitable framework.

5.2.3 Clarification Reading

In order to handle CRs suitably, a dialogue system should be able to distinguish between, say, clausal and constituent CR readings where possible. It should be noted that this may not always be possible – the corpus analysis performed revealed examples where CPs mistook one reading for another: in example (23), the anonymous interviewer's initial constituent CR is mistaken for a clausal "check" and must be repeated in a less ambiguous (non-reprise) manner.

(23)[2]
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

It seems plausible that constituent CRs will generally be produced when clarifying some object or word that has not been encountered previously in the dialogue. If this is the case, the dialogue IS could be used to maintain a list of constituents that can no longer be queried, helping to disambiguate further CRs. A corpus search is proposed to provide evidence to support or disprove this.

Intonation might also be a clue to clausal/constituent disambiguation – Grice et al. (1995) report that the L+H tone is present in information-seeking queries (like constituent CRs) but is not required in confirmation-seeking checks (some clausal readings). Similarly G&S claim that *echo* and *reference* reprise interrogatives are intonationally distinct, although they assign both a clausal reading.

[2] BNC file H5G, sentences 193–196

5.2.4 Clarification Referent

It seems likely that information state would be the best source of disambiguation information for this problem. In example (11) (repeated here as example (24)), the reprise sluice *"Who?"* cannot be taken as clarifying *Leon* in the source utterance (as Leon is one of the CPs), so is taken as clarifying *she*:

(24)[3]
Sarah: Leon, Leon, sorry she's taken.
Leon: **Who?**
Sarah: Cath Long, she's spoken for.

5.2.5 Elliptical Nature

Both intonation and IS might be useful in determining the correct elliptical resolution of a fragment. While there appears to be a "typical" interrogative clause intonation in English (rising boundary tone?) and similar for declaratives (falling or steady boundary tone?), there may also be a distinctive reprise intonation. Some evidence for this is presented by Grice et al. (1995).

This may differ between fragments and sluices (as interrogative intonation differs between wh- and polar questions), and possibly between reprise forms and readings (see above), but any distinction between reprise and other forms would help to disambiguate examples such as example (21).

5.2.6 Combining Sources of Information

Interplay between several different sources of information such as a probabilistic parser and (possibly probabilistic) constraints from intonation and IS poses the problem of how to combine such information in a coherent fashion.

While rule-based approaches to such problems are an option, they tend to get very complicated very quickly – a simpler and more robust approach might be a probabilistic one such as the use of Bayesian nets.

[3] BNC file KPL, sentences 347–349

Bibliography

Peter Bohlin (Ljunglöf), Johan Bos, Staffan Larsson, Ian Lewin, Colin Matheson, and David Milward. Survey of existing interactive systems. In *Task Oriented Instructional Dialogue (TRINDI): Deliverable 1.3*. University of Gothenburg, 1999a.

Peter Bohlin (Ljunglöf), Robin Cooper, Elisabet Engdahl, and Staffan Larsson. Information states and dialogue move engines. In Jan Alexandersson, editor, *IJCAI-99 Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, 1999b.

Manuela Boros, Maria Aretoulaki, Florian Gallwitz, Elmar Nöth, and Heinrich Niemann. Semantic processing of out-of-vocabulary words in a spoken dialogue system. In *Proceedings of the European Conference on Speech Communication and Technology*, volume 4, pages 1887–1890, 1997.

Johan Bos. Predicate logic unplugged. In *Proceedings of the Tenth Amsterdam Colloquium*. ILLC/Department of Philosophy, University of Amsterdam, 1995.

Holly P. Branigan, Martin J. Pickering, and Alexandra A. Cleland. Syntactic co-ordination in dialogue. *Cognition*, 75:13–25, 2000.

Chris Brew. Stochastic HPSG. In *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics*, pages 83–89. University College, Dublin, 1995.

Lou Burnard. *Reference Guide for the British National Corpus (World Edition)*. Oxford University Computing Services, 2000. URL `ftp://sable.ox.ac.uk/pub/ota/BNC/`.

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

Herbert H. Clark. *Using Language*. Cambridge University Press, 1996.

Iakovos Dallas. An HPSG-based travel agent dialogue system. Master's thesis, King's College, London, forthcoming.

Tony Dodd. *SARA: Technical Manual.* Oxford University Computing Services, 1997. URL `http://info.ox.ac.uk/bnc/sara/TechMan/`.

Gregor Erbach. Prolog with features, inheritance and templates. In *Proceedings of the Seventh Conference of the European Association for Computational Linguistics*, pages 180–187, 1995.

Raquel Fernández Rovira. SHARDS: It's great. Technical report, King's College, London, forthcoming.

Charles Fletcher. Levels of representation in memory for discourse. In Morton Ann Gernsbacher, editor, *Handbook of Psycholinguistics.* Academic Press, 1994.

Florian Gallwitz, Elmar Nöth, and Heinrich Niemann. A category based approach for recognition of out-of-vocabulary words. In *International Conference on Spoken Language Processing*, volume 1, pages 228–231, 1996.

Simon Garrod and Martin Pickering. Toward a mechanistic psychology of dialogue: The interactive alignment model. In P. Kühnlein, H. Rieser, and H. Zeevat, editors, *Proceedings of the Fifth Workshop on Formal Semantics and Pragmatics of Dialogue.* BI-DIALOG, 2001.

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

Jonathan Ginzburg and Robin Cooper. Resolving ellipsis in clarification. In *ACL/EACL01 Conference Proceedings.* Association for Computational Linguistics, July 2001.

Jonathan Ginzburg, Howard Gregory, and Shalom Lappin. SHARDS: Fragment resolution in dialogue. In Harry Bunt, Ielka van der Sluis, and Elias Thijsse, editors, *Proceedings of the Fourth International Workshop on Computational Semantics (IWCS-4)*, pages 156–172. ITK, Tilburg University, Tilburg, 2001a.

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

Jonathan Ginzburg, Ivan A. Sag, and Matthew Purver. Integrating conversational move types in the grammar of conversation. In P. Kühnlein, H. Rieser, and H. Zeevat, editors, *Proceedings of the Fifth Workshop on Formal Semantics and Pragmatics of Dialogue.* BI-DIALOG, 2001b.

Martine Grice, Ralf Benzmüller, Michelina Savino, and Bistra Andreeva. The intonation of queries and checks across languages: Data from map task dialogues. In *Proc. XIII International Congress of Phonetic Sciences, Stockholm*, 1995.

Irvine L. Hetherington. *A Characterization of the Problem of New, Out-of-Vocabulary Words in Continuous-Speech Recognition and Understanding.* PhD thesis, Massachusetts Institute of Technology, 1994.

Albert S. Hornby. *Oxford Advanced Learner's Dictionary of Current English.* Oxford University Press, third edition, 1974. With the assistance of Anthony P. Cowie and J. Windsor Lewis.

Jacqueline Kowtko, Stephen Isard, and Gwyneth Doherty. Conversational games within dialogue. In *Proceedings of the ESPRIT Workshop on Discourse Coherence*, 1991.

Staffan Larsson, Peter Ljunglöf, Robin Cooper, Elisabet Engdahl, and Stina Ericsson. GoDiS - an accommodating dialogue system. In *Proceedings of ANLP/NAACL-2000 Workshop on Conversational Systems*, 2000.

Oliver Lemon, Anne Bracy, Alexander Gruenstein, and Stanley Peters. Information states in a multi-modal dialogue system for human-robot conversation. In P. Kühnlein, H. Rieser, and H. Zeevat, editors, *Proceedings of the Fifth Workshop on Formal Semantics and Pragmatics of Dialogue.* BI-DIALOG, 2001.

Ian Lewin and Stephen Pulman. Inference in the resolution of ellipsis. In *Proceedings of the ESCA Workshop on Spoken Dialogue Systems*, pages 53–56, 1995.

Bernd Ludwig. Dialogue understanding in dynamic domains. In P. Kühnlein, H. Rieser, and H. Zeevat, editors, *Proceedings of the Fifth Workshop on Formal Semantics and Pragmatics of Dialogue.* BI-DIALOG, 2001.

George A. Miller. WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41, November 1995.

David Milward. Dynamic dependency grammar. *Linguistics and Philosophy*, pages 561–605, 1994.

David Milward. Towards a robust semantics for dialogue using flat structures. In *Amstelogue '99, Workshop on the Semantics and Pragmatics of Dialogue*, volume II. University of Amsterdam, 1999.

David Milward. Distributing representation for robust interpretation of dialogue utterances. In *Proceedings of the 38th ACL Conference.* MIT Press, 2000.

Roger Mitton. *Oxford Advanced Learner's Dictionary of Current English: expanded "computer usable" version.* Oxford Text Archive, 1992. URL `http://ota.ahds.ac.uk/`.

Massimo Poesio. Semantic ambiguity and perceived ambiguity. In Kees van Deemter and Stanley Peters, editors, *Semantic Ambiguity and Underspecification*, number 55 in CSLI Lecture Notes. CSLI Publications, 1996.

Matthew Purver. SCoRE: A tool for searching the BNC. Technical report, Department of Computer Science, King's College London, 2001.

Matthew Purver, Jonathan Ginzburg, and Patrick Healey. On the means for clarification in dialogue. In *Proceedings of the 2nd ACL SIGdial Workshop on Discourse and Dialogue.* Association for Computational Linguistics, September 2001.

Uwe Reyle. Co-indexing labeled DRSs to represent and reason with ambiguities. In Kees van Deemter and Stanley Peters, editors, *Semantic Ambiguity and Underspecification*, number 55 in CSLI Lecture Notes. CSLI Publications, 1996.

Frank Richter. *A Mathematical Formalism for Linguistic Theories with an Application in Head-Driven Phrase Structure Grammar.* PhD thesis, Eberhard-Karls-Universität Tübingen, 2000. Version of April 28th, 2000.

Frank Richter and Manfred Sailer. Underspecified semantics in HPSG. In Harry Bunt and Reinhard Muskens, editors, *Computing Meaning*, pages 95–112. Kluwer Academic Publishers, 1999.

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

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

Paul Taylor, Simon King, Stephen Isard, Helen Wright, and Jaqueline Kowtko. Using intonation to constrain language models in speech recognition. In *Proc. Eurospeech '97*, pages 2763–2766, Rhodes, Greece, 1997.

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

Teun A. van Dijk and Walter Kintsch. *Strategies of Discourse Comprehension.* Academic Press, 1983.

Gertjan van Noord, Gosse Bouma, Rob Koeling, and Mark-Jan Nederhof. Robust grammatical analysis for spoken dialogue systems. *Natural Language Engineering*, 1(1):1–48, 1998.

Karsten L. Worm and C. J. Rupp. Towards robust understanding of speech by combination of partial analyses. In *European Conference on Artificial Intelligence*, pages 190–194, 1998.

Appendix A

Schedule

Period	SHARDS	Clarifications	Underspecification	Disambiguation
July 2001 – September 2001		Corpus Results Confirmation Experiment Design New CR Readings	Unknown Words	
October 2001 – December 2001	Dialogue Management	Experiments New CR Readings/Forms	Unknown Words	Corpus Search
January 2002 – March 2002	Grammar Expansion	Analysis of Experiments New CR Readings/Forms Implementation	Unknown Referents	Corpus Search Analysis
April 2002 – June 2002			Ambiguity	Stochastic Grammar Information State
July 2002 – September 2002			Ambiguity	Information State Intonation
October 2002 – December 2002			Ambiguity	Intonation
January 2003 – March 2003				Intonation
April 2003 – September 2003	Write up thesis			

Table A.1: Proposed Work Schedule