Higher Order Unification and the Interpretation of Focus

Stephen G. Pulman

SRI International Cambridge Computer Science Research Centre [1]
and University of Cambridge Computer Laboratory.

November 1995

Contents

[1] SRI International, Suite 23, Millers Yard, Mill Lane, Cambridge CB2 1RQ, UK.
Email: pulman@cam.sri.com

Abstract

Higher order unification is a way of combining information (or equivalently, solving equations) expressed as terms of a typed higher order logic. A suitably restricted form of the notion has been used as a simple and perspicuous basis for the resolution of the meaning of elliptical expressions and for the interpretation of some non-compositional types of comparative construction also involving ellipsis. This paper explores another area of application for this concept in the interpretation of sentences containing intonationally marked 'focus', or various semantic constructs which are sensitive to focus.

Similarities and differences between this approach, and theories using 'alternative semantics', 'structured meanings', or flexible categorial grammars, are described. The paper argues that that the higher order unification approach offers descriptive advantages over these alternatives, as well as the practical advantage of being capable of fairly direct computational implementation.

1 Introduction

Higher order unification (Huet 1975) is a way of combining information (or equivalently, solving equations) expressed as terms of a typed higher order logic. A suitably restricted form of the notion has been used as a simple and perspicuous basis for the resolution of the meaning of elliptical expressions (Dalrymple, Shieber and Pereira 1991) and for the interpretation of some non-compositional types of comparative construction also involving ellipsis (Pulman 1991; Alshawi and Pulman 1992; Gawron 1992, 1995). This paper explores another area of application for this concept in the interpretation of sentences containing intonationally marked 'focus', or various semantic constructs which are sensitive to focus.

The structure of the paper is as follows. In the remainder of this section, we describe some basic observations concerning the semantic and pragmatic effects of focus interacting with a range of different linguistic phenomena, and describe and compare three current approaches to the interpretation of focus: these theories are from Krifka (1991, 1992); Steedman (1994); and Rooth (1985,1992).

In section 2 we describe the notion of higher order unification, as used by Dalrymple, Shieber and Pereira in their analysis of ellipsis. We extend the use of higher order unification to focus phenomena, and present an explicit grammar fragment which accurately covers the basic cases.

In section 3 we sketch how to extend the fragment to deal with adverbs and other focus phenomena, and discuss some interactions between focus and context which the present theory seems to deal with quite elegantly. The section, and the paper, concludes with a brief illustration of how the theory might be used to construct a component of a system generating sentences with appropriate focus markings.

1.1 The range of focus phenomena

The term 'focus' has had many meanings in linguistics and computational linguistics (see for example, Hajičová 1984, Sgall 1984; Brown and Yule 1983, Sidner 1983, Carter 1987). The way I shall use the term here is essentially as it is used in the work of the Prague School, in Selkirk 1984 (who curiously omits to acknowledge any debt to the Prague school, in an otherwise comprehensive and scholarly work), and in Rooth 1985, 1992; Krifka 1992, etc. I assume that focus is a property of semantic rather than ofsyntactic units, and is signalled variously by particular syntactic constructs (for example, clefts, topicalisation, etc), or by pitch accent. It is customary to distinguish 'broad' and 'narrow' focus, roughly according to the size of the constituent, as illustrated by the following question-answer pairs. Again as is customary, the location of the focus is indicated by upper case.

1 What did she do? She ATE THE APPLE. *(broad, on VP)*
2 Did she drop it? No, she ATE it. *(narrow, on V)*

Broad focus is often to do with the raising of salience of some information by either introducing it or emphasising it. Narrow focus often introduces a contrastive element in which a previous assertion is denied or called into question in some way. However, there is no necessary connection between the scope of focus and the information function conveyed by it, and the converse correspondences are often found. Also, I am being deliberately non-committal in describing these pragmatic effects because none of the traditional notions of 'topic-comment', 'assertion-presupposition' or 'given-new' seem wholly accurate in describing the effects of focus, although they have all been used to do so (see e.g. Brown and Yule, (op cit), Delin (1990) for some pertinent observations).

In what follows, I shall usually concentrate on 'narrow' focus phenomena, since intuitions about the scope of the focussed elements are much clearer than for the broad focus examples. However, the account proposed works equally well for the broad cases: it is just more difficult to identify this unambiguously. (For some suggestions, see Selkirk, 1984, 206-251).

Consider some simple examples where the context makes the identification of the focus artificially clear:

3	a	A: Does John like Mary?	B: No, John likes JANET
	b	A: Does John fear Mary?	B: No, John LIKES Mary
	c	A: Does Janet like Mary?	B: I don't know, but JOHN likes Mary
	d	A: Who likes who?	B: JOHN likes JANET

From these examples, it is easy to see how terms like 'given-new' or 'presupposed-asserted' come to be used: intuitively, the second member of the pair can be seen as presenting its information in two parts: one part is already predictable from the context; the second part is an addition to it and is what is contrasted with some element occurring in the 'antecedent'. The semantic (as opposed to the pragmatic) problem of the interpretation of focus is how to derive these two pieces of information in as compositional a way as possible from the syntactic and semantic structure, and the focus marking if any, of the sentence.

The range of further focus-sensitive phenomena is well-known. There are focus particles like 'only', 'even', 'too' and many others; quantifying determiners like 'most'; quantificational adverbs like 'always'; manner adverbials like 'quickly' (although their sensitivity to focus has been largely ignored in the focus literature); negation; generics; aspectual composition, and no doubt many others. The following list of examples (from Rooth, 1985; Krifka, 1991; 1992; and elsewhere) illustrates some of the basic facts that any theory of the interpretation of focus must handle.

4 John only introduced Bill to SUE
5 John only introduced BILL to Sue
6 John only introduced BILL to SUE
7 John even drank WATER
8 Even$_1$ John$_1$ drank only$_2$ water$_2$
9 John even$_1$ [only$_2$ [drank water]$_2$]$_1$
10 John even$_1$ drank [only$_2$]$_1$ [water]$_2$

Examples 4 to 6 show that truth conditions may vary with focus; and that complex foci are possible with focus particles. Example 7 shows the basic case of 'even'; the remaining examples show that two focus particles can occur together in a sentence with separate, embedded, or identical foci respectively. The focus associations are indicated by subscripts.

The truth conditions associated with quantificational adverbs are also affected by focus:

11 Mary always took JOHN to the movies
12 Mary always took John to the MOVIES
13 LONDONERS most often go to Brighton
14 Londoners most often go to BRIGHTON

Example 11 says that whenever Mary took someone to the movies, it was John, as opposed to 12, which says that whenever she took John somewhere, it was to the movies. The remaining pair also differ in truth conditions: the first says that most visitors to Brighton are from London, whereas the second says that the typical holiday destination for a Londoner is Brighton. The second is compatible with more than half of the visitors to Brighton being from places other than London, whereas the first is not.

Some determiners are focus-sensitive (Krifka 1990:509):

15 Most ships pass through the LOCK at night
16 Most ships pass through the lock at NIGHT

It is easiest to tease apart the difference in truth conditions with reference to the broad or neutral focus interpretation of the sentence. With this focus, the sentence says that the number of ships going through the lock at night is greater than the number of ships going by any other route (e.g. the canal) at any other time. But 15 is consistent with most ships, taken overall, going through the canal during the day. It will be true provided only that at night there are fewer ships going through the canal etc. than through the lock. Example 16 is also consistent with most ships, taken overall, going through the canal during the day, and will be true provided only that more ships go through the lock at night than during the day.

Many other phenomena are focus-sensitive: but these are the basic cases that any theory of focus interpretation must be able to describe adequately.

1.2 Rooth's 'Alternative Semantics'

The most influential current theory of focus is perhaps that advanced by Rooth (1985,1992), dubbed 'alternative semantics'. Two notions are important in this approach: the semantic value of a constituent, which is just the ordinary denotation that some Montague-style semantics would assign to it, and the 'focus' semantic value. The focus semantic value of a constituent containing a focus is the set of meanings obtained by substituting all available entities of the appropriate type for the focussed phrase. Thus if a sentence has a semantic value representable as 'A(b,c)', where c is focussed, the focus semantic value will be:

17 $\{ \mathrm{A(b,x)} \mid \mathrm{x} \in \mathrm{Y}\}$

where Y is the domain of things of the type of c.

Focus is interpreted via a 2 place operator $\sim$ which is adjoined to some constituent C containing a focus, F. The containing constituent is the first argument of $\sim$ and a free variable, V, is the second:

18 $[_C \text{ ... F ...}]{\sim}\mathrm{V}$

This variable is constrained to be of the type either of members of the focus semantic value for C, or sets of such. The interpretation of C $\sim$ V is that

> (i) the expression asserts C, and presupposes that
>
> (iia) V is a (pragmatically determined) subset of the focus semantic value of C containing both the semantic value for C and another distinct member (set case), OR
>
> (iib) that V is a member of the focus semantic value of C distinct from the semantic value (individual case).

V is interpreted by being associated with some contextually available antecedent meaning: it is essentially interpreted as an anaphor.

In interpreting a sentence like

19 $[_S\ [_{NP}$ John$]\ [_{VP}$ likes $[_F$ Janet$]]]{\sim}$V

(with focus indicated in Rooth's notation), in a context provided by the question 'Who does John like?', things proceed as follows. The free variable V, in this context, will be instantiated to the semantic value of the preceding question, which we assume to be {like(john,x) | x $\in$ type e}. Then the expression asserts that 'like(john,janet)' and presupposes that {like(john,x) | x $\in$ type e} is a subset of the focus semantic value of the assertion, which it is (though not a proper subset, since it happens in this case to be identical).

In the case where the assertion was made in a different context: e.g.

20 A: John likes Jane. B. No, John likes JANET

then the free variable V would be instantiated to 'like(john,jane)' and the presupposition would be that this is a member of the focus value of the assertion distinct from the semantic value itself. The presupposition is satisfied, since 'likes(john,jane)' is an instance of 'likes(john,x)' but distinct from 'likes(john,janet)'.

In Rooth's account, an example like:

21 John only $[_{VP}$introduced BILL to Sue$]$ $\sim$ V

has, for the VP containing the focus, focus and semantic values F and S:

22 F = { λx. introduce(x,y,Sue) | y of type e }
S = λx. introduce(x,bill,sue)

The sentence, by virtue of the meaning of 'only' presupposes that John introduced Bill to Sue (i.e. S is true of John), and asserts that no member of F other than S is true of John. By virtue of the interpretation of focus it also presupposes that there is a contextually available value for V which is a member of F distinct from S.

Rooth's theory has several desirable properties. Focus is given a uniform interpretation wherever it occurs. No construction specific information has to be associated with the interpretation of the $\sim$ operator, at least where truth conditions are not affected. In cases where focus affects truth conditions, some qualification has to be made to this claim. Rooth discusses 'only' (p111) and presumably the same qualification would extend to other adverbial or determiner cases.

1.3 Structured Meanings

Building on work by Jacobs (1983), von Stechow (1989) and others, Krifka (1991, 1992) has articulated a structured meaning approach to the semantics of focus-sensitive constructions. Focus induces a partition of a meaning into a background and focus component, represented $\langle$B,F$\rangle$, where the original meaning is equivalent to an application B(F). This is achieved by wholly compositional means. Focus on a constituent with a semantic representation A introduces a focus-background structure of the form $\langle\lambda$X.X, A$\rangle$, where the X in the identity function (the 'empty background') has the same type as A. The focus-background structure is further built up by composing the applications specified by semantic rules with the background part of the structure. Thus if the semantic rule specifies that B should be applied to A (where A is not focussed) then the focus equivalent of the rule will specify that B applied to $\langle\lambda$X.X, A$\rangle$ is $\langle\lambda$X.B(X),A$\rangle$, or more generally, that B applied to $\langle\lambda$X.C, D$\rangle$ is $\langle\lambda$X.B(C),D$\rangle$. If the application in the original rule was A(B), where A is focussed, then the new application of $\langle\lambda$X.X, A$\rangle$(B) is defined as $\langle\lambda$X.X(B), A$\rangle$; more generally: $\langle\lambda$X.C,D$\rangle$(B) = $\langle\lambda$X.C(B),D$\rangle$.

We can illustrate this process using a simplified version of Krifka's fragment.

1.	S	$\rightarrow$	NP VP	: NP(VP)
2.	VP	$\rightarrow$	V_t NP	: V_t(NP)
3.	VP	$\rightarrow$	V_i NP to NP	: V_i(NP1)(NP2)
4.	C	$\rightarrow$	C_F	: $\langle\lambda$X.X, C$\rangle$ (i.e. focus any constituent)
5.	C	$\rightarrow$	FO C	: FO($\langle$*background,focus*$\rangle$)
6.	NP	$\rightarrow$	Name	: λP.P(Name)
7.	V_t	$\rightarrow$	like, etc.	: λO.λx.O(λy.like(x,y))
8.	V_i	$\rightarrow$	introduce	: λO.λI.λx.I(λz.O(λy.introduce(x,y,z)))

Unless otherwise specified, variables *x, y, z* are of type e; *P, Q, R* of type $\langle$e,t$\rangle$, and *S, O, I* of type $\langle\langle$e,t$\rangle$,t$\rangle$.

In rule 5, the meaning for the daughter C must be a $\langle$background,focus$\rangle$ pair: FO is a focus particle like 'only'.

The interpretation of an example like:

23 John likes MARY

will proceed as follows. Following Krifka (1991:33), we assume an abstract illocutionary focus particle 'assert' is introduced by applying rule 5 to the otherwise complete sentence meaning, as detailed below.

Rule	Phrase	Meaning
6	Mary	$= \lambda$P.P(m)
5	MARY	$= \langle\lambda$X.X, λP.P(m)$\rangle$
2	likes MARY	$= \langle\lambda$X.[λO.λz.O(λy.like(z,y)](X), λP.P(m)$\rangle$
		$= \langle\lambda$X.λz.X(λy.like(z,y)), λP.P(m)$\rangle$
6	John	$= \lambda$Q.Q(j)
1	John likes MARY	$= \langle\lambda$X.[λQ.Q(j)](λz.X(λy.like(z,y))), λP.P(m)$\rangle$
		$= \langle\lambda$X.X(λy.like(j,y))), λP.P(m)$\rangle$
5		= assert($\langle\lambda$X.X(λy.like(j,y)), λP.P(m)$\rangle$)

Krifka's definition of 'assert($\langle$B,F$\rangle$)' (p33-34) entails, roughly, that the final meaning representation asserts that the background B is true of the focus F (i.e. 'like(j,m)'), subject to the felicity condition that there are other parallel or comparable alternatives to the focus which the background could have been informatively predicated of. (We shall see later that this is not quite complete. It is also necessary in some circumstances that there are alternatives to the background, too).

Now for the derivation of:

24 John only introduced BILL to Sue

6	Bill	$= \lambda$P.P(b)
4	BILL	$= \langle\lambda$X.X, λP.P(b)$\rangle$
6	Sue	$= \lambda$Q.Q(s)
8	introduce	$= \lambda$O.λI.λx.I(λz.O(λy.intr(x,y,z))) $= V$
3	introduce BILL to Sue	$=$ [$\langle\lambda$X.V(X),λP.P(b)$\rangle$](λQ.Q(s))
		$= \langle\lambda$X.[V(X)](λQ.Q(s)),λP.P(b)$\rangle$
		$= \langle\lambda$X.λx.X(λy.intr(x,y,s)), λP.P(b)$\rangle$
5	only introduce BILL to Sue	= only($\langle\lambda$X.λx.X(λy.intr(x,y,s)), λP.P(b)$\rangle$

The functor 'only' is defined:

25 only($\langle$B,F$\rangle$) $\leftrightarrow$ λx.[B(F)(x) & $\forall$y.[parallel(y,F) & B(y)(x) $\rightarrow$ y = F]]

The predicate 'parallel' I use to mean 'an alternative' (Rooth), 'parallel' (DSP) or 'comparable' (Krifka). When the definition of 'only' is spelled out we get as the meaning of the VP:

26 λx. intr(x,b,s)
& $\forall$Y.[parallel(Y,λP.P(b))
& [λX.λz.X(λy.intr(z,y,s))](Y)(x) $\rightarrow$ Y = λP.P(b)]

This reduces to:

27 λx. intr(x,b,s)
& $\forall$Y.[parallel(Y,λP.P(b))
& Y(λy.intr(x,y,s)) $\rightarrow$ Y = λP.P(b)]

When the subject and the VP are combined this is the property that will be predicated of John. It can be paraphrased (ignoring the fact that the NP meanings have quantifier types) as saying that every person that John introduced to Sue was Bill: i.e. tjhere was no other.

Krifka extends this treatment to cases like:

28 John only introduced BILL to SUE
29 JOHN likes MARY

by augmenting the logic with a list-forming operator, with associated types and rules of application. Krifka distinguishes cases like this, where one focus operator is associated with two foci, as 'complex focus', as opposed to 'multiple focus', where we have more than one operator-focus pair. We will examine all these more complicated examples in some detail later on, for they present a challenging test-bed for any analysis of the mechanisms involved in the interpretation of focus.

In Krifka 1992 the analysis is extended to focus-sensitive determiners and adverbials, and in Krifka 1993 to various anaphoric and presuppositional phenomena involving focus.

1.4 Combinatory Categorial grammar

In a series of papers, Steedman (1990, 1991, 1993) has presented a theory of focus which exploits (and is partly intended to justify) the flexibility of his version of categorial grammar. In this theory, words and phrases can be assembled into constituents in many semantically equivalent ways, leading to what is often called the 'spurious ambiguity' problem. Steedman argues that at least some of these derivations are not spurious, but correspond to differences of information structure and are reliably associated with different intonational properties.

Steedman discusses information structure in terms of 'theme' vs. 'rheme', reserving the term 'focus' for the intonational nucleus of the theme or rheme, either of which can be further decomposed into a focus-ground structure. We can illustrate the main features of the approach with these examples (Steedman 1993:18):

30 What about MUSICALS? Who admires THEM?
31 [$_{RHEME}$ MARY] [$_{THEME}$ admires MUSICALS]
32 What about MARY? What does SHE admire?
33 [$_{THEME}$ MARY admires] [$_{RHEME}$ MUSICALS]

Steedman accounts for the differences in interpretation by associating them with different syntactic derivations appropriate to the different intonation patterns, as indicated. The major syntactic constituents along with the intonational contour associated with them determine the information structure of the sentence.

The two relevant derivations are as follows, where T indicates type raising, A is application and C is composition:

```
Mary        admires      musicals
NP          (S\NP)/NP     NP
------->T
S/(S\NP)
------------------->C
      S/NP
      --------------------->A
               S

Mary        admires      musicals
NP          (S\NP)/NP     NP
            ----------------->A
                  S/NP
------->T
S/(S\NP)
----------------------->A
          S
```

Within each theme and rheme there can be a further focus-ground subdivision, at least in more complicated examples:

34 a Mary envies the man who wrote the musical.
But who does she admire?

b [$_{THEME}$ Mary ADMIRES]
[$_{RHEME}$ the woman who DIRECTED the musical]

The lower case portions of the theme and rheme are background, and the upper case the focus components.

Using this flexibility of the constituent structures permitted by combinationary categorial grammar, Steedman is able to provide analyses of the basic focus phenomena which, as he points out, deliver interpretations basically identical to those obtained by Rooth and Krifka. The advantage of Steedman's approach is claimed to be that, contrary to these other theories, no extra combinatory mechanisms specific to focus phenomena are required.

1.5 Comparisons

It is clear that these three approaches to the interpretation of focus have a great deal in common, in particular in the kind of interpretation for focus sentences that they provide. Von Stechow (1991) explicitly compares the Alternative Semantics and Structured Meanings approaches. He shows that many of the analyses of Rooth's 1985 theory can all be recast in the Structured Meaning framework. However, he points out that the reverse is not quite the case, in that Rooth's theory gives the wrong results for some cases where the set of alternatives generated by the constituent containing focus needs to be further contextually restricted:

35 a Did Sir John introduce each gentleman to his partner?
b No, Sir John only introduced BILL to MARY

Clearly example (b) should not be false if Sir John has in fact introduced people other than Bill and Mary who are not ⟨gentleman,partner⟩ pairs. A mechanism like that employed for question-answer pairs achieves the correct effect within the Structured Meaning analysis. (In Rooth's 1992 version of the theory this omission is remedied (p 79) where it is assumed that some pragmatic process of restriction of domain of quantification is possible in such cases.)

Krifka (1992:218) points to another disadvantage of Alternative Semantics (which in this case also holds of the later version) as opposed to Structured Meanings. Notice that in Rooth's theory, the meanings and presuppositions of sentences containing focus are stated purely in terms of the 'focus semantic value' (the value of a constituent containing a focus) and the usual semantic value of the expressions. The focus semantic value (from which the alternatives are obtained by a pragmatic restriction) does not explicitly mention, or allow access to, the meaning of the focussed item itself.

In Krifka's view, focus with quantifying adverbials frequently leads to an exhaustive interpretation:

36 Mary always took JOHN to the movies

i.e. she took John and no-one else. But the kind of representation[1] that is obtained for the meaning of this sentence on Rooth's theory, acording to Krifka, is

37 $\{s \mid \text{took}(m,j,s)\ \&\ \forall s'.[\exists x.\ x \text{ in ALT}(j)\ \&\ \text{took}(m,x,s') \rightarrow \text{took}(m,j,s)]\}$

The antecedent of the conjunction, however, is compatible with situations in which Mary took more than one individual to the movies, provided that she also took John. In order to obtain the exhaustive interpretation it is required to specify that 'j' is the only member of his alternatives for which the antecedent is

[1] In Krifka's paper (p219) the example for which this is the representation contains no adverbial. I assume this is an error.

satisfied. But of course that in turn requires us to be able to have direct access to the meaning of the focussed item 'John'. The structured meaning theory gives us this access via the structuring of the sentence into background and focus. But the 'alternative semantics' account only gives us access to the focus semantic value of the consituent containing the focus item. Krifka goes on to develop an account on which such examples are semantically ambiguous between an exhaustive and a non-exhaustive reading (p230).

Both the Alternative Semantics and the Structured Meaning approaches are founded on the assumption that focus is a property of syntactic constituents (Rooth 1992: 76; Krifka 1991:17). Thus as Krifka acknowledges (1991:35,47), there is a problem with cases in which focus appears on something other than a constituent. There are two types of problem case: the first of these is the case where focus appears on a constituent, but that constituent is a discontinuous one:

38 a Er hat [sich] nur [RASIERT]
He has self only shaved = he has only shaved himself
b John only [turned] it [OFF]

Focus can be on 'sich rasiert' and 'turned off'. In a theory at which multiple levels of representation are available it would be plausible to argue that the focus was on a constituent at some 'deep' level. However, Krifka's structured meaning theory is a surface-oriented one and is thus unable to handle these cases properly (1991:17-18).

The second type of case concerns focus on a sequence of items that do not form a constituent at any orthodox level of syntactic or semantic representation. For example, in the following question-answer pair (Krifka 1991:47) it is natural to interpret focus as falling on a predicate meaning 'was kissed by Sue':

39 a What happened to John?
b SUE KISSED him

However, since the structured meaning approach builds on the usual compositional structure of sentence meanings, it will not be possible to achieve the appropriate separation into focus and background for such examples. Krifka suggests that a categorial grammar approach might be needed; his own alternative is to analyse such examples as topic-comment effects, claiming that 'the relevant examples of purported non-constituent focus all have a purported background that IS a constituent' (p 47). The 'background' in examples like these constitute the 'topic' and the focus is the 'comment'. Krifka suggests (p 47) that topic-comment structures can be derived using structured meanings.

Unfortunately for this superficially plausible approach, it seems clear that the process of constructing structured meanings will run into exactly analogous problems to the non-constituent focus cases. We have already seen illustrations of non-constituent topics (=theme) in the examples used by Steedman, repeated here:

40 What about MARY? What does SHE admire?
41 [$_{THEME}$ MARY admires] [$_{RHEME}$ MUSICALS]

Structured meanings will be just as difficult to construct for non-constituent backgrounds as for non-constituent focus.

It is clear that in this area, Steedman's combinatory categorial grammar treatment of focus has a distinct descriptive advantage over Alternative Semantics or Structured Meanings. The availability of alternative non-standard derivations makes it possible to produce the right kind of constituent structure for these examples. (Note however that Steedman actually assumes that focus is a property of words, not of constituents 1993:19). The charge that these structures are not otherwise necessary can be quite convincingly rebutted by pointing to various apparent examples of non-constituent coordination that also seem to need these non-standard constituent structures:

42 Mary admires but Susan detests musicals

The analysis in terms of non-standard constituents seems at least as convincing as analyses involving deletion or movement.

Unfortunately, there are other reasons to doubt whether the categorial approach is the correct one to take to the analysis of focus. One immediate problem for such an account, raised by Joshi (1990), is that the constituent structure needed to capture the effects of focus may conflict with that needed to capture the syntactic facts. Consider a sequence like:

43 What about MARY? What does SHE admire?
44 B: Mary admires MUSICALS but detests opera.

In order for the focus structure to be captured, the first conjunct has to have a derivation of the form '(Mary admires) musicals', as we have seen. But in order for the VP conjunction to be well formed, the first conjunct needs to have the structure 'Mary (admires musicals)'. Categorial grammar provides many derivations for a sentence but has not so far required two different ones simultaneously.

A further problem arises from the isomorphism between syntactic and semantic units in categorial grammar. In this framework, syntactic constituents are semantic units too. There is a wider variety of both than in standard theories of grammar, but the close linking means that it is to be expected *a priori* that Steedman's approach will have problems with those cases that have been used to argue that focus is not syntactically based, like Rooth's example:

45 They only asked whether I knew the woman
who chairs the ZONING board

In this example focus is on something within a syntactic island, making it impossible to construct the required component of the meaning by syntactic means, either by movement in a standard theory, or by a series of type-raisings and compositions in categorial grammar. (Island constraints are enforced in Steedman's theory of grammar by mechanisms described in Steedman 1987). The predicate that is required for Rooth's example is:

46 λx. they asked whether I knew the woman who chairs x

Steedman's response to this dilemma is that there are two factors involved in constructing this component of the meaning in focus constructions, since, as we saw earlier, both theme and rheme can themselves be decomposed into focus and background. The theme supplies the open proposition proper. The background of the rheme 'constrains the set(s) of alternative entries quantified over' (25). Thus the open proposition is:

47 λx. they asked whether I knew x

where x is restricted to a contextually restricted set whose members satisfy

48 λy. the woman who chairs y

where the syntactic correlates of both of these meanings ARE permitted by Steedman's grammar.

However, it seems to me that Steedman's position on island constraints is not altogether consistent here. The syntactic configuration needed to permit the structure of the open proposition in example 47 is itself usually regarded in the literature as an island violation. And the predicate in 48 which is required to constraint the domain of quantification is an instance of the same island violation as in the original example. (Steedman only describes the predicate informally, but it must have something like this structure). If Steedman's grammatical theory permits such open propositions to be generated it will also allow for the generation of examples with analogous structure like:

49 *What does he know whether John likes?
50 *Who does he know what irritates?

- or, indeed, the original example. At the very least, some further argumentation is necessary as to why these island violations are sometimes to be permitted, while other violations of the constraints illustrated by Rooth's example are supposed to be impossible.

A further difficulty for Steedman would appear to be presented by those cases in which two focussed constituents appear non-contiguously, either as complex foci, or as a discontinuous constituent:

51 John only introduced MARY to SUE
52 John only TURNED it OFF

In order to construct the right semantic entities, the theory would need to be able to regard these examples as having, somehow, the following structures:

53 [John only introduced to][MARY] [SUE]
54 John only [TURNED OFF] [it]

While categorial grammar can provide non-standard constituents constructed of adjacent sequences of words using only the standard mechanisms of semantic assembly, more complex 'wrapping' operations will be necessary to provide the type of derivation needed here. However, the general availability of such operations would greatly increase the 'spurious ambiguity' problem that flexible categorial grammar already has. Notice that such operations would also compromise the 'incrementality of interpretation' that is often claimed to be a virtue of categorial grammars.

2 A higher order unification based analysis

2.1 Higher Order Unification

In this section we describe the background to our own treatment of focus, describing higher order unification and illustrating its use in building semantic representations.

Higher order unification (Huet 1975) is defined on pairs of higher order terms of a typed lambda calculus. These terms have to be in a normal form, in which the types of functional terms are made explicit via lambda variables. The precise details of this normal form, and of the unification algorithm itself are rather too complex for brief presentation here. However, a few illustrations should give a sufficient feel for the concept.

In higher order logic, variables of any type are allowed, not just those ranging over individuals. Thus whereas

55 P(john)=sleeps(X) $\rightarrow$ {P=sleeps,X=john}

is not a well-formed first order equation, it is a well formed higher order one, and with the solution given. The substitution representing the solution to the equation will, when applied to each of the input expressions, make them identical. Some more complex examples are:

56	a	P(john)=likes(john,mary)	$\rightarrow$ {P=λx.likes(x,mary)}
	b	P(x)=likes(john,mary)	$\rightarrow$ {P=λy.likes(y,mary),x=john}
			$\rightarrow$ {P=λy.likes(john,y),x=mary}

While the notion of a unifying substitution is almost the same as for the first order case, the notion of identity is not: if a unifying substitution is found, application of the substitution to the expressions may not produce identical terms until beta-reduction has been carried out. Notice that, unlike the first order case, there is no 'most general' unifier: for the second example there are two different solutions, neither being noticably less general than the other. In fact there will usually be an infinite number of solutions for any given case, although most of them will be only trivial variants of the interesting one. For example, even given our first simple pair of terms

57 P(john)=sleeps(X) $\rightarrow$ {P=sleeps,X=john}

we ignored the alternative solutions of the form

58 {P=λy.sleeps(y),X=john},{P=λy.[λz.sleeps(z)](y),X=john} etc.

Thus higher order unification is not deterministic: there may be multiple solutions. It is also the case that higher order unification is not decidable: if there is no solution, an algorithm implementing higher order unification may not terminate. [2]

The attraction of higher order unification is that it provides a clear logical (and computational) interpretation of one particular notion of 'combination of information' in a way which is much less tied to the precise syntax of the expressions involved than other types of information combining operations, such as those based on pattern matching, or indeed, those using first order unification.

To illustrate the use of higher order unification, and the particular approach to interpretation taken here, we shall use the example of ellipsis. The analysis is derived from that presented by Dalrymple, Shieber and Pereira (1991), (henceforth DSP).

I will make several assumptions about the mechanisms involving in building up meaning representations. Firstly, I assume that the syntax-driven meaning composition process produces 'quasi-logical forms' (QLF) (cf. Alshawi and Crouch 1992) which, when they contain contextually dependent constructs, require further contextual processing to determine their truth-conditions. Ellipsis focus phenomena are instances (among many others) of such constructs, and their interpretation at the QLF stage will be represented by various functors ('ellipsis', 'even', 'only' etc.). These functors will be interpreted via 'conditional equivalences' (Rayner and Alshawi, 1992; Rayner 1993) which specify what the meaning

[2]From the examples just given, it is intuitively easy to see how the algorithm could get stuck in an infinite recursion, trying ever larger and larger terms as candidate hypotheses. A practical implementation could largely overcome these problems by (a) returning only the first N solutions and (b) imposing a time-out, regarding time-out as failure. There are several experimental systems based on higher order unification which just ignore the undecidability altogether. See Miller and Nadathur 1986 for a description of λ-Prolog, a higher order version of Prolog, and for some interesting examples of its application.

of such a construct will be, if various contextual conditions hold or can be consistently assumed. We assume a setting in which relevant aspects of the context are represented by expressions in some logic: checking contextual conditions amounts to seeing whether they hold in, or can consistently be added to the context.

We shall represent at the QLF level the meaning of constructions containing ellipsis by means of a polymorphic functor 'ellipsis' applying to the meanings of the elements actually present in the construction. The type of the functor is that of $\langle A_1....A_n,t\rangle$ where $A_1....A_n$ are the types of the constituents present, sometimes called the 'remnants'. (A fully general treatment for multiple constituents requires the type system of our logic to have, as well as the usual type constructor for functional types, one which constructs tuples or sequences of types. We shall suppress this kind of detail in what follows, in the interests of readability, and represent multiple arguments directly, or using the notation A•B•C•... to denote lists of arguments.)

The meaning of the 'ellipsis' functor is given by a conditional equivalence of roughly the following form:

$$\begin{array}{l} \text{ellipsis}(A_1\bullet...\bullet A_n) \Leftrightarrow Q(A_1\bullet...\bullet A_n) \\ \quad \text{if} \\ \qquad \text{antecedent}(C), \\ \qquad \&\ Q(B_1\bullet...\bullet B_n) = C, \\ \qquad \&\ \text{parallel}(A_1\bullet...\bullet A_n,\ B_1\bullet...\bullet B_n). \end{array}$$

The first condition requires us to locate an appropriate antecedent logical form, C, which is salient in the context. (This may be a previous utterance or an earlier part of the current utterance, depending on the construction). The second condition sets up a higher order equation which has the effect of decomposing the antecedent into a predicate Q applied to a sequence of one or more elements $B_1...B_n$ which the third condition requires to be pairwise parallel to the elements in the elliptical fragment. To be parallel, two terms need to be at least of the same type and have the same sortal properties. But in general, deciding what counts as being parallel may involve a variety of morphological, syntactic, semantic and discourse factors: for some discussion see DSP, p 406-7, Prüst 1992, and below.

Notice that the approach taken here differs slightly from that in DSP in that a higher order equation *simultaneously* generates candidates for parallel items as well as for the ellipsis resolution itself. In DSP location of parallel elements is a separate stage not involving unification: however, as we shall see, location of parallel elements actually requires similar machinery, and for similar reasons.

The effect of satisfying these conditions is to find an instantiation for Q which, when substituted in, provides the interpretation of the ellipsis on the right hand side of the equivalence. As mentioned above, there may be many solutions to a higher order equation, some of which are trivial or inappropriate in some way. Conditions prohibiting vacuous abstraction, and some further conditions to ensure that only sensible solutions[3] are generated are required. We shall come to some of these later.

Take the following example of ellipsis, using simplified logical forms:

59	a	A: John likes Mary	LF= likes(john,mary)
	b	B: ... and Bill (too)	QLF = ellipsis(bill))

Using the conditional equivalence to interpret the QLF 'ellipsis(bill)' instantiates A to the single remnant 'bill', and the antecedent, C, to 'likes(john,mary)' and sets up a higher order equation of the form:

60 Q(B) = likes(john,mary)

There are two solutions to this equation that are of interest, since in this case either 'john' or 'mary' can be the value of B, as they each count as parallel to 'bill'. On one solution, Q = 'λz.likes(z,mary)', and on the other Q = 'λz.likes(john,z)'. Thus the right hand side of the equivalence will be one of the following expressions:

([λz.likes(z,mary)](bill)
= likes(bill,mary)
([λz.likes(john,z)](bill)
= likes(john,bill)

Both of these are possible interpretations of the sentence. Notice that in examples like these, it would not be possible to achieve the same results by a process of copying of either syntactic or semantic constituents, for the simple reason that for one of these readings there is, on any standard[4] theory of syntactic and semantic structure, no such constituent available. Neither 'John likes', nor 'λz.likes(John,z)' would be formed by compositional syntactic or semantic rules.

As remarked above, in this treatment of ellipsis, unlike that in DSP, we have conflated the process of solving the relevant equation with that of finding the appropriate parallel elements, using higher order unification to generate the candidate parallel elements as part of the solution to the equation. This is necessary, since just as the antecedent logical form may not contain the resolvent as a

[3]Note that this need not be a practical problem: an implementation of higher order unification can be designed so as to enumerate the sensible before the silly solutions most of the time. In the Huet algorithm, for example, ordering the operation of 'projection' before that of 'imitation' suffices to postpone many of the trivial solutions.

[4]Some versions of categorial grammar may make such constituents available. However, as DSP point out(fn. 3, 403), there are grounds for suspicion that such theories are in fact assuming something like higher order unification, but in disguise.

constituent, so there is no guarantee that it will contain constituents of the appropriate type to act as parallel elements. This can be again illustrated by the previous example: the meaning for the elliptical sentence 'and Bill' is represented in our simplified QLF as 'ellipsis(bill)'. However, 'Bill' is actually a full NP in this context, and thus a more accurate logical form would be one that assigned to it the usual type of NPs: 'λP.P(bill)'. This would make the variable Q, in this instance, of type $\langle\langle\langle e,t\rangle,t\rangle,t\rangle$. Thus a slightly more realistic formulation of the partially solved equations would be:

$$Q(\lambda P.P(john)) = likes(john,mary)$$
$$Q(\lambda P.P(mary)) = likes(john,mary)$$

This gives as solutions for Q the more complex expressions:

$$\lambda R.R(\lambda z.likes(john,z))$$
$$\lambda R.R(\lambda z.likes(z,mary))$$

It is clear that no simple process of substitution will allow us to retrieve the appropriate parallel elements from the antecedent clause, since they no longer exist in the appropriate form after beta-reduction. Equally, no simple process of substitution will allow us to construct an appropriate candidate for Q here. But we can make higher order unification do both jobs simultaneously, generating candidates both for parallel elements and instantiations for ellipsis variables.

2.2 An illustrative fragment

In this section I give a fragment which will provide appropriate meaning representations for some basic focus phenomena, using higher order unification to provide interpretations for QLFs. The fragment is based on that presented by Krifka (1991:24 ff) and covers approximately the range of data discussed there, most of which was described in an earlier section.

I assume that we can identify which elements of the sentence meaning are focussed, either because of their syntactic configuration, or via intonation, (although this assumption will be relaxed later), and that these elements are available as arguments to the QLF functor representing the contribution of the focus construct. Focussed elements need not be constituents: in the limit, a sequence of meanings of individual morphemes is a valid argument to a focus functor. (Various mechanisms, within a unification-based syntactic and semantic formalism, can be envisaged to accomplish this 'association with focus'.) For simplicity, I will ignore issues of interaction with quantifier scope, pronominals, etc.

Here are some simplified rules:

0.	S	$\rightarrow$ S	:	assert(*Foci*, S))
1.	S	$\rightarrow$ NP VP	:	NP(VP)
2.	VP	$\rightarrow$ V_t NP	:	V_t(NP)
3.	VP	$\rightarrow$ V_i NP_o NP_{io}	:	$V_i(NP_o)(NP_{io})$
4.	VP	$\rightarrow$ FAdv VP	:	λx.FAdv(*Foci*,VP(x))
5.	NP	$\rightarrow$ FAdv NP	:	λP.FAdv(*Foci*,NP(P))
6.	NP	$\rightarrow$ Name	:	λP.P(Name)
7.	V_t	$\rightarrow$ kiss	:	λO.λx.O(λy.kiss(x,y))
8.	V_i	$\rightarrow$ introduce	:	λO.λI.λx.I(λz.O(λy.introduce(x,y,z)))

Unless otherwise specified, variables *x, y, z* are of type e; *P, Q, R* of type $\langle$e,t$\rangle$, and *S, O, I* of type $\langle\langle$e,t$\rangle$,t$\rangle$.

Rule 0 introduces an abstract QLF illocutionary operator (following Jacobs and Krifka). The arguments to 'assert' are the meaning of the focused element(s) of the sentence, and the meaning of the S constituent itself.

Rules 4 and 5 introduce the focus sensitive adverbs 'only' and 'even' as modifiers of VP and NP respectively. The meanings associated with these words introduce a QLF construct 'only' or 'even' which has arguments analogous to those of 'assert', except that the second argument of the focus functor is regarded as polymorphic, and has the type required for the constituent in question. The remaining rules are self-explanatory.

We also show the conditional equivalences which give a contextual definition of the various focus operators. In general, we follow Krifka's assumptions about the meaning of these terms, with some minor simplifications. In stating these equivalences we use two predicates which are taken as primitives (although in reality each of them encapsulates some notoriously slippery notions): 'context(X)', which means that X is or becomes contextually salient; and 'parallel(X,Y)' which means that X and Y are the same sort of thing ('comparable' is the term used by Krifka). Since we frequently need the notion of things being parallel but not identical, we also use the infix symbol 'X$\approx$Y' to mean 'parallel(X,Y) & X$\neq$ Y'. As we shall see, in the case where 'parallel' takes several arguments - e.g. parallel(X•Y,A•B), meaning 'parallel(X,A) & parallel(Y,B)'- there seems to be a pragmatic condition that at least one parallel pair are distinct: this is a generalisation of Krifka's felicity condition on asserting (1991:34). The symbol '=' is elsewhere interpreted as higher order unification.

Firstly we give an equivalence which provides an interpretation for the functor 'assert':

assert(F, S) $\Leftrightarrow$ S
  if

$$
\begin{aligned}
&\mathrm{B(F) = S}\\
&\text{\& context(C)}\\
&\text{\& P(A) = C}\\
&\text{\& parallel(B}\bullet\text{F, P}\bullet\text{A)}
\end{aligned}
$$

This says that asserting S with focus F is equivalent to S if S can be structured into a background B and a focus F; where some proposition C is salient, and can also be structured into components P and A such that P and B, and A and F, are parallel. In such a context the felicity condition that the background could have been informatively predicated of something other than the focus will be met.

We give an equivalence to interpret the functor 'only':

$$
\begin{aligned}
&\text{only(F,S)} \Leftrightarrow \text{no X. X}\approx\text{F \& B(X)}\\
&\quad \text{if}\\
&\qquad \text{context(S)}\\
&\qquad \text{\& B(F) = S.}
\end{aligned}
$$

In giving this definition for 'only' we depart slightly from the analysis provided by Krifka 1991, in that the basic proposition (S) conveyed is presupposed rather than asserted. The equivalence says that 'only(F,S)' holds if S is contextually salient (or is made so by accommodation) and if S can be structured into a background-focus pair: the interpretation is that there does not exist any X parallel to F but different from it, for which the background predicate B holds. (This component is basically equivalent to the analysis familiar from Rooth and other authors, but given in a less verbose form.)

Finally, an equivalence giving the interpretation of 'even':

$$
\begin{aligned}
&\text{even(F,S)} \Leftrightarrow \text{S}\\
&\quad \text{if}\\
&\qquad \text{B(F) = S}\\
&\qquad \text{\& }\exists\text{X. X}\approx\text{F \& B(X)} \rangle \text{ B(F)}
\end{aligned}
$$

This analysis of 'even' is based on Krifka's, and uses an operator $\rangle$ which can be interpreted as a kind of subjective probability. An expression like 'A $\rangle$ B' means something like 'A is believed more probable than B'. The equivalence says that 'even(F,S)' is equivalent to S if S can be structured into B and F, and there is something parallel but not identical to F which B would be more likely to be true of.

To illustrate the mechanisms we have introduced let us begin with a simple narrow focus example:

61 Did John kiss Sue?
John kissed MARY

Rule	phrase:	meaning:	focus
7	MARY	λP.P(m)	λP.P(m)
2	kissed MARY	[λO.λx.O(λy.kiss(x,y))](λP.P(m))	λP.P(m)
		= λx. kiss(x,m)	λP.P(m)
7	John	λQ.Q(j)	
1	John kissed MARY	[λQ.Q(j)](λx.kiss(x,m))	λP.P(m)
		= kiss(j,m)	λP.P(m)
0	John kissed MARY	assert(λP.P(m), kiss(j,m))	

The equivalence for 'assert' requires us to locate a suitable contextually salient proposition: 'kiss(john,sue)' and solve the following higher order equations:

62 B(λP.P(m)) = kiss(j,m)
P(A) = kiss(j,s)

The solution we need is that B = P = λO.O(λx.kiss(j,x)), with A = λQ.Q(s), where the quantifier meanings corresponding to 'Mary' and 'Sue' are, we assume, parallel in the required sense (but distinct as required by the pragmatic condition). The meaning of the sentence can be informally glossed as 'John kissed Mary (as opposed to Sue, Jane, etc)'.

Before moving on to cases of complex and multiple focus, let us first see how the higher order unification analysis solves the problem of non-constituent focus. Consider an example like:

63 What happened to Mary?
JOHN KISSED her

(We will simplify by assuming that the pronoun has been resolved.)

7	JOHN	λP.P(j)	λP.P(j)
	KISSED Mary	λx.kiss(x,m)	λO.λy.O(λz.kiss(y,z))
1	J KISSED M	kiss(j,m)	[λP.P(j)]$\bullet$[λO.λy.O(λz.kiss(y,z))]
			= Foci
0	J KISSED M	assert(Foci, kiss(j,m))	

We have to first solve an equation of the form B(F1$\bullet$F2):

64 B([λP.P(j)]$\bullet$[λO.λy.O(λz.kiss(y,z))]) = kiss(j,m)

This gives us the solution that B = λS.λV.S(V(λQ.Q(m))), where V is of the type of transitive verbs. We assume that the context variable C is instantiated to the previous question represented as '$\exists$x.happen(x,m)'. So we now need to solve a second equation:

65 $\quad P(A1\bullet A2) = \exists x.\text{happen}(x,m)$

$$\begin{aligned} P &= \lambda S.\lambda V.S(V(\lambda Q.Q(m)) \\ A1 &= \lambda P.\exists x.P(x) \\ A2 &= \lambda O.\lambda y.O(\lambda z.\text{happen}(y,z)) \end{aligned}$$

where we assume the solution meets the required conditions on parallelism. The meaning can be glossed as 'Mary had the property of being kissed by John, as opposed to ...'.

Thus we are able to accommodate cases of non-constituent focus (or of course, of focus on discontinuous constituents) without the need of further mechanisms. The only assumption we need to make is that when focus is indicated on sequences of words that do not form a constituent, the meanings of that sequence of words are available to the focus functor in question: this is surely the weakest assumption possible under these circumstances.

We now turn to cases of complex focus: that is, where one focus particle is associated with two independent focussed constituents.

66 John only introduced BILL to SUE

3	introduced BILL to SUE	: λx.int(x,b,s)	λP.P(b)•λQ.Q(s) = Foci
4	only introd B to S	: λx.only(Foci,int(x,b,s))	
1	John only introd B to S	: only(Foci,int(j,b,s))	

Rule 0 is not applicable, in the intended reading, because the focussed constituents have already been associated with a focus functor. The equivalence for 'only' requires us to solve the equation:

67 $\quad B(\lambda P.P(b)\bullet\lambda Q.Q(s)) = \text{int}(j,b,s)$

$\quad B = \lambda O.\lambda I.O(\lambda x.I(\lambda y.\text{int}(j,x,y)))$

The equivalence presupposes that 'int(j,b,s)', and says that:

68 $\quad \text{only}(\lambda P.P(b)\bullet\lambda Q.Q(s),\text{int}(j,b,s)) \Leftrightarrow$

$\quad \text{no } X\bullet Y.\ X\bullet Y\approx[\lambda P.P(b)]\bullet[\lambda Q.Q(s)]$

$\quad \&\ [\lambda O.\lambda I.O(\lambda x.I(\lambda y.\text{int}(j,x,y)))](X\bullet Y)$

Paraphrasing this (and abstracting away from the quantifier types involved) we get that John introduced no other X to Y than Bill to Sue.

Next we turn the the case of multiple focus particles, each with an independent associated focus, indicated by subscripts:

69 $\quad$ Even$_1$ John$_1$ drank only$_2$ water$_2$

We will simplify the exposition by treating 'water' as if it were a name: nothing crucial hangs on this.

6	water	:	λP.P(w)	
5	only water	:	λQ.only(λP.P(w),Q(w))	λP.P(w)
2	drank only water	:	λx.only(λP.P(w),drink(x,w))	
5	even John	:	λP.even(λR.R(j),P(j))	λR.R(j)
1	even J drank only w	:	even(λR.R(j),only(λP.P(w),drink(j,w)))	

Using the equivalence for 'even' we have that the expression is equivalent to 'only(λP.P(w),drink(j,w))' under the conditions stipulated. We have first to solve the equation, which has the solution indicated:

70 B(λR.R(j)) = only(λP.P(w),drink(j,w))
B = λS.only(λP.P(w),S(λx.drink(x,w)))

Then the presupposition associated with 'even' is:

71 $\exists$X.X$\approx\lambda$R.R(j) &
[λS.only(λP.P(w),S(λx.drink(x,w)))](X) $\rangle$ only(λP.P(w),drink(j,w))

This says that there is something parallel to (the meaning for) John of which it is more plausible to predicate the property of drinking only water.

The meaning 'only(λP.P(w),drink(j,w))' will, by the application of the equivalence for 'only', presuppose that John drank water, and assert that:

72 no X. X$\approx$[λP.P(w)] & [λO.O(λx.drink(j,x))](X)

i.e. there is nothing parallel but not identical to water such that John drank it.

Next we have the case of nested focus structures. In an example like:

73 John even_1 [only_2 [drank water]$_2$]$_1$

the focus associations are as indicated. The meaning derivation is:

2	drank water :	λx.drink(x,w)
4	only drank water :	λy.only(λx.drink(x,w),drink(y,w)) = VP
4	even only drank water :	λz.even(VP,only(λx.drink(x,w),drink(z,w)))

The meaning for the whole sentence is then:

74 even(λy.only(λx.drink(x,w),drink(y,w)), only(λx.drink(x,w),drink(j,w)))

The equivalence for 'even' tells us that the sentence therefore means

75 only(λx.drink(x,w),drink(j,w)))

subject to the presupposition determined as follows:

76 B(λy.only(λx.drink(x,w),drink(y,w))) = only(λx.drink(x,w),drink(j,w))
B = λP.P(j)

Substituting and reducing we get:

77 $\exists X.X{\approx}[\lambda y.\mathrm{only}(\lambda x.\mathrm{drink}(x,w),\mathrm{drink}(y,w))]$
$\& \ X(j) \rangle \mathrm{only}(\lambda x.\mathrm{drink}(x,w),\mathrm{drink}(j,w))$

In other words, the presupposition is that there is at least one property comparable to only drinking water which is more likely to be a property of John.

Obtaining the meaning of '$\mathrm{only}(\lambda x.\mathrm{drink}(x,w),\mathrm{drink}(j,w))$' requires us to solve another equation:

78 $B(\lambda x.\mathrm{drink}(x,w))=\mathrm{drink}(j,w)$
$B = \lambda P.P(j)$

After substituting and reducing in the right hand side of the equivalence we obtain '$\mathrm{no}\ X.\ X{\approx}[\lambda x.\mathrm{drink}(x,w)] \ \&\ X(j)$', which is the correct interpretation: no relevant property other than that of drinking water holds of John.

Krifka also points to examples where the same focus can be associated with two different particles:

79 John even$_1$ only$_2$ drank [[water]$_2$]$_1$

In his and our framework this requires focus elements to continue to be available for other focus operators even after they have already been associated with one. This seems to be generally possible, as for example when contrastive focus is used along with a a focus particle:

80 A: John drank only water.
B: No, John drank only WINE

Here 'wine' must be available as the focus for 'only' as well as for the 'assert' operator.

The derivation for example 79 is:

5	only drank water :	$\lambda x.\mathrm{only}(\lambda P.P(w),\mathrm{drink}(x,w)) = \mathit{VP}$
5	even only drank water :	$\lambda z.\mathrm{even}(\mathit{VP}, \mathrm{only}(\lambda P.P(w),\mathrm{drink}(z,w)))$
1	John even only drank water :	$\mathrm{even}(\mathit{VP}, \mathrm{only}(\lambda P.P(w),\mathrm{drink}(j,w)))$

Solving the equation for 'even' we have:

81 $B(\lambda x.\mathrm{only}(\lambda P.P(w),\mathrm{drink}(x,w))) = \mathrm{only}(\lambda P.P(w),\mathrm{drink}(j,w))$
$B = \lambda P.P(j)$

giving a presupposition:

82 $\exists X.X{\approx}[\lambda x.\mathrm{only}(\lambda P.P(w),\mathrm{drink}(x,w))]$
$\&\ [X(j) \rangle \mathrm{only}(\lambda P.P(w),\mathrm{drink}(j,w))]$

Solving the equation for 'only' (see above for the same equation) gives us:

83 $\mathrm{only}(\lambda P.P(w),\mathrm{drink}(j,w))$

$= \text{no X.X}\approx[\lambda\text{P.P(w)}]\ \&\ [\lambda\text{O.O}(\lambda\text{x.drink(j,x))}](\text{X})$

and we arrive at the final interpretation that example 79 presupposes both that John drank water, and that some comparable but different property than drinking only water is more likely to be true of John, and the sentence asserts that there is nothing other than water that was drunk by John.

Finally we come to the case where a focus particle is itself a focus.

84 John even_1 drank $[\text{only}_2]_1$ $[\text{water}]_2$

6	water	λQ.Q(w)	
5	only water	λP.only(λQ.Q(w),P(w))	only
2	drank only water	λx.only(λQ.Q(w),drink(x,w))	only
5	even drank only water	λz.even(only, only(λQ.Q(w),drink(z,w))	

In interpreting 'even(only, only(λQ.Q(w),drink(j,w)))' we solve the equation:

85 B(only) = only(λQ.Q(w),drink(j,w))
B = λO.O(λQ.Q(w),drink(j,w))

giving us the presupposition:

86 $\exists$X. X$\approx$only & X(λQ.Q(w),drink(j,w)) $\rangle$ only(λQ.Q(w),drink(j,w))

This can be paraphrased as saying that it is more likely that, say, John ALSO drank water than that John ONLY drank water (cf. Krifka 1991:33).

The remaining aspects of the interpretation of this sentence are already familiar.

Finally, let us note that we have an analogue of the problem of 'primary' vs. 'secondary' occurrence of terms discussed in DSP (p404). They point out that secondary occurrences of terms (which typically arise from resolving pronouns or reflexives) can only be abstracted over in solving an equation if the primary term on which they are dependent is abstracted over too. If this constraint was not imposed then their analysis of ellipsis would predict the following reading, among others, for this example:

87 Joe admires himself, and Bill does too
admire(joe,joe) & P(bill)
P(joe) = admire(joe,joe) $\rightarrow$ P =λx.admire(joe,x)
P(bill) = admire(joe,bill)

The solution in c is inadmissible because only the secondary occurrence of 'bill' has been abstracted over.

Now consider the following examples:

88 a Only John likes John
b John only likes JOHN

These examples show that we must be able to identify which occurrence of the term 'John' is associated with the focus operator when solving the higher order equation, and only abstract over that, for otherwise both sentences will be three ways ambiguous, corresponding to solutions in which the background is one of:

89 1 λx.likes(x,x)
2 λx.likes(John,x)
3 λx.likes(x,John)

The appropriate instantiation for example 88a is 3, and for example b it is 2, which is what the proposed restriction to focus will ensure. This is a different restriction from the primary/secondary restriction proposed for ellipsis by DSP, and where they are in conflict the restriction to focus seems to take precedence.

90 a Only John likes himself
b John only likes HIMSELF

In example a, with no contrastive focus, the only available interpretation is that no-one other than John has the property 'λx.likes(x,x)'. In order to obtain this reading the occurrence of John that is associated with the focus particle 'only', and that recovered from the reflexive (by the primary/secondary principle), must both be abstracted over. In example b, on the other hand, the interpretation is that no-one other than John has the property 'λx.like(John,x)', in apparent violation of the primary/secondary principle. The difference is that in this latter case, 'himself' is the focussed constituent associated with 'only', and although dependent on 'John', it is not 'John' that is focussed.

3 Some extensions

3.1 Adverbs

As discussed earlier, a variety of adverbials are focus sensitive. We have already mentioned quantificational adverbs: some other types are:

91 Surprisingly, MARY likes John
92 Surprisingly, Mary likes JOHN
93 I accidentally threw away your BOOK
94 I accidentally THREW AWAY your book

What is surprising is different in examples 91 and 92, just as what is accidental is different in the following pair of examples

Even simple manner adverbials display this sensitivity to focus:

95 John swam the HELLESPONT quickly
96 John SWAM the Hellespont quickly

‘Quickly’, like the corresponding adjective, requires an implicit standard. In the first example, (in the intended interpretation), John’s swim is quick by the standards of swimmings by John. In the second, John’s swim is quick by the standards of all relevant types of crossings by John of the Hellespont. It would be possible for the second to be true (if his journey by raft and canoe were slow) even under the circumstances that the Hellespont was the slowest swim John completed (making the first false).

The higher order unification analysis of focus can be extended to examples like these straightforwardly. I will illustrate with respect to manner and quantificational adverbs only: extension to the other types is routine. We assume an analysis of sentences in which the basic tenseless meaning of the sentence is a predicate on events: for the ‘quickly’ examples it would be:

97 λe. swim(e,john,hellespont)

The meaning of the VPs of such sentences (ignoring adverbs) will be of the form ‘λe.λx.swim(e,x,hellespont)’ and the sentence rule will combine the subject NP and the VP by composition: ‘λe.NP(VP(e))’ when the VP is of such a form. Tense and aspect, in a fuller treatment, would be operators on the resulting predicate (and the context) producing the fully contextually resolved interpretation for the sentence. We will ignore their contribution, to keep the quasi-logical forms simple.

Given such an analysis, we can assume that manner adverbs are introduced by a rule something like:

98 VP $\rightarrow$ VP MAdv (or MAdv VP) : madv(Focus,MAdv,VP)

VPs will now be of type $\langle$e,$\langle$e,t$\rangle\rangle$. In the following examples, event or situation variables will be *e, f, g*, other variables of type e will be *a, b, c, x, y, z.* Expressions like ‘λxyx...’ are to be interpreted as ‘curried’, i.e. abbreviations for ‘λx. λy. λz....’

The QLF construct ‘madv’ is interpreted by an equivalence:

$$\begin{aligned}&\text{madv(F,MAdv,VP)} \Leftrightarrow \lambda\text{ex.[VP(e)(x) \& MAdv(e,}\lambda\text{f.}\exists\text{y.B(y)(f)(x))]}\\ &\quad\text{if}\\ &\qquad\text{B(F) = S.}\end{aligned}$$

The effect of solving the higher order equation is to abstract over the focus constituent, and construct a ‘standard’ by existentially quantifying over that position. We assume that the domain of the quantifier is pragmatically restricted to items parallel to the focus.

The equivalence will be instantiated as follows for example 95:

QLF $= \text{madv(h,quick,}\lambda\text{ga.swim(g,a,h))}$
B(h) $= \lambda\text{ga.swim(g,a,h))}$
B $= \lambda\text{bga.swim(g,a,b)}$

The interpretation given by the right hand side of the equivalence will be:

99 $\lambda\text{ex.}[\lambda\text{ga.swim(g,a,h)}](\text{e})(\text{x}) \;\&\; \text{quick}(\text{e},\lambda\text{f.}\exists\text{z.}[\lambda\text{bga.swim(g,a,b)}](\text{z})(\text{f})(\text{x}))$
$= \lambda\text{ex.swim(e,x,h)} \;\&\; \text{quick}(\text{e},\lambda\text{f.}\exists\text{z.swim(f,x,z)})$

When combined with the subject John, this will say that the event was quick by the standards of events of John swimming something.

Example 97 will result in a right hand side of:

100 $= \lambda\text{ex.swim(e,x,h)} \;\&\; \text{quick}(\text{e},\lambda\text{f.}\exists\text{Z.Z(f,x,h)})$

where Z is a variable of the type of 'swim', and when applied to the subject it will say that the event was quick by the standards of relevant actions concerning John and the Hellespont.

Let us turn to quantificational adverbs. I shall analyse them in more or less the same way as Rooth (1985, 1995) and Krifka (1992), that is, as generalised quantifiers over events or situations. A generalised quantifier is a relation between a predicate determining a 'restriction' and one determining a 'body' set. A schema for the interpretation of temporal adverbs would thus be:

adverb(restriction,body)

The 'body' of the quantifier is the basic event predication of the sentence containing the adverb. The restriction will be an expression describing the contextually appropriate restriction set of events: what Schubert and Pelletier (1989) call the 'reference ensemble'. The adverb itself will be a relation between the size of the restriction set and the 'intersection set' of the restriction and the body.

The difficult part of this analysis for previous treatments has been how to determine a suitable expression to play the role of the restriction. It will always involve some non-trivial contextual reasoning. The QLF for sentences involving these quantificational adverbs will therefore have a construct for the restriction indicating that it has to be resolved with respect to the current context. In some cases, the expression may be resolved from immediate linguistic context:

101 When he visits Mary, John always takes flowers

This will give us something like:

102 $\text{all}(\lambda\text{e.visit(e,john,mary)},\lambda\text{f.take(f,john,flowers)}$

This says that the size of the set of events of visiting by John of Mary, is the same size as the set of events that are simultaneously visitings of Mary by John, and takings of flowers by John. Alternatively, every event that is a visit of Mary by John is also a taking of flowers by John. This can only be the case if there are no events of John visiting Mary that are not also events of John taking flowers, which is the desired result. (The paraphrase may sound more natural if 'events' are replaced by 'situations').

In other cases, a combination of linguistic and non-linguistic resources have to be used:

103 (In a context in which John is expected at the restaurant)
John's always late

This should be interpreted as something like:

104 all(λe.visit(e,john,restaurant),λf.late(f,john)

The cases we are interested in are those in which there is narrow focus on some constituent in the sentence. The focus item can be used to determine the restriction set (as originally proposed in Rooth 1985), on the basis of a generalisation of the body set.

To keep our story simple, we shall assume that quantificational adverbs are interpreted at a level where the whole sentence meaning has been assembled, and that their QLF representation, and associated equivalence is of the form:

qadv(F,QAdv,S) $\Leftrightarrow$ QAdv(Restr, S)
if
Restr is contextually determined, or
B(F) = S, and Restr = λe.$\exists$f.B(f)(e)

The restriction set describes the 'contextually relevant' set of events: the 'reference ensemble'. This is either determined from the context, or, if there is no salient set of events in the context, by solving a higher order equation. The restriction set is constructed by quantifying over the position of the focussed element in building up the meaning, giving a description of a (usually) bigger set than that described by the body, S. The relation between the restriction set and the intersection set captures the quantificational force of the adverb.

Now we can return to our examples: solving the equation and substituting into the right hand side of the equivalence gives us:

105 Sue always takes JOHN to the movies
QLF = qadv(j,always,λe.take(e,s,j,m))
B(F) = λe.take(e,s,j,m) $\rightarrow$ B = λze.take(e,s,z,m)
rhs = always(λe.$\exists$a.[λzf.take(f,s,z,m)](a)(e), λe.take(e,s,j,m))
= always(λe.$\exists$a.take(e,s,a,m),λe.take(e,s,j,m))

We can paraphrase this meaning as saying that 'Every event of taking by Sue of someone to the movies, is a taking by Sue of John to the movies'.

With a different focus, we get a different meaning:

106 Sue always takes John to the MOVIES
always(λx.$\exists$a.take(x,s,j,a),λe.take(e,s,j,m))

Krifka (1992:218) points out with respect to example 105 that the 'non-exhaustive' interpretation, in which it is not excluded that Sue also took people other than John, is not the only one for such sentences. The sentence can also be understood as saying that whenever Sue took someone, it was always John and no-one else. This interpretation is one that motivates the requirement for the focus meaning to be directly accessible, distinguishing the Structured Meaning theory from Alternative Semantics.

Krifka proposes two different logical forms, one effectively including the meaning of 'only', to capture this distinction. However, I think this is not a genuine ambiguity, but arises from the contextual restriction on the range of the existential quantifier in the restriction. If the quantifier is understood to be ranging only over singular individuals then we get the exhaustive interpretation. If it ranges over plural individuals, then by making a reasonable assumption about the behaviour of distributive predicates, we can instead obtain the non-exhaustive interpretation.

If this is not a correct diagnosis, nothing prevents us from adapting Krifka's analysis within our framework, since we also have direct access to the meaning of the focused item.

The remaining types of adverb mentioned earlier may not depend on focus for their truth conditions in the way that manner and quantificational adverbs do. Adverbs like 'surprisingly' and 'accidentally' are probably to be analysed as involving a propositional attitude paraphraseable, somewhat clumsily, as 'it is surprising to me that ..' and 'it was accidental that I ....'. It is a matter of some debate as to whether focus affects the truth conditions of propositional attitudes (see for example, Dretske 1979). One relevant factor is presumably how fine grained a theory of propositions is in play. Whether the difference is one of presupposition or of truth conditions, it should be straightforward to extend the current approach to construct the appropriate interpretations. It is also straightforward to analyse focus effects in determiners like 'most', or in negation, or elsewhere by using higher order equations to construct appropriate representations as we have done in the examples given so far.

3.2 Focus and Context

In giving a conditional equivalence to interpret the 'assert' operator, we required that some relevant salient proposition was available in the context. Is it plausible to suppose that there is always some such proposition available in the context? If fact, given our assumption that conditions can be abduced if they cannot be deductively proven, it is not necessary to assume that the contextual proposition is already available. If an appropriate candidate can be reconstructed the context can be augmented by 'accommodating' or 'abducing' the relevant proposition.

Given the further fact that we are using higher order unification, which is a completely non-directional information sharing operation, the possibility arises that the context might actually be determined by the interpretation of focus. In fact, the same non-directionality suggests that it may be possible for focus itself to be determined by the context. Interpreting focus involves the interplay of three elements: a context; the content of an utterance; and the focus marking of the utterance. Clearly, the content of the utterance must always be specified, but the logical properties of the mechanisms we are using suggests that it is theoretically possible for the focus to be unspecified, but constrained by the context: or the reverse.

It is clearly the case that given an unambiguous content and focus marking, a competent native speaker can reconstruct a context that would make the utterance felicitous, and our approach provides an account of that ability without any modification. Consider the situation in which you overhear, with no prior context, the following snatch[5] of conversation:

107 ... so I bought another boat and THAT sank

The most obvious context that would motivate such a use of narrow focus is one in which the first boat also sank, although of course the sentence does not say this explicitly.

Processing the first conjunct gives us the information that there are two boats in the context: 'boat1', and 'boat2', which is the discourse referent we construct for 'another boat'. We infer the existence of 'boat1' from the meaning of 'another'. 'Boat2' is the one the speaker bought: he may or may not have also bought the first one. We process the second conjunct (after interpretation of 'that') using the relevant equivalence, repeated here:

assert(F, S) ⇔ S
 if
  B(F) = S
  & context(C)
  & P(A) = C
  & parallel(B•F, P•A).

S is 'sank(boat2)' and the variables B and F in this equivalence can be instantiated

[5]The example is from a talk by Stephen Isard, many years ago.

to ‘sank’ and ‘boat2’ respectively but we are, by hypothesis, unable to find a candidate for C.

Recall the status of the conditions in a conditional equivalence. We have assumed that they are checked deductively against the current context (and whatever background knowledge may be necessary). It is appropriate (and intended) that each condition should be regarded as something like a theorem to be proved with respect to the ‘axioms’ representing the context. As with familiar theorem proving methods, or programming languages like Prolog, a side effect of this may be to find instantiations for ‘query variables’ in the theorem. Thus we can treat a condition like ‘parallel(B•F, P•A)’ as a call to find instantiations of any uninstantiated variables which meet the parallelism condition.

In the current example the call will be:

108 parallel(sank•boat2, P•A)

Possible instantiations for A will be ‘boat1’, and a possible instantiation for P will be the value of B, i.e. ‘sank’. (Identifying A with ‘boat2’ would yield a context that could not be informatively updated by this utterance). These instantiations jointly satisfy the pragmatic requirement on parallelism and are available from the context supplied by processing the content of the whole sentence.

```
boat(boat1)
boat(boat2)
bought(I,boat2)
```

The values for P and A of course now allow us to infer that the context, C, should be ‘sank(boat1)’ and as this is consistent the context can be augmented to accommodate this proposition. This inference of a plausible context seems like a faithful reconstruction of what happens in the most salient interpretation of this sentence.

If, on the other hand, the snatch we overheard had been:

109 ... so I bought another boat and that SANK

then the value of the variable F would be the predicate ‘sank’, and B would have been resolved by the equation to ‘λQ.Q(boat2)’. This time we have a call of:

110 parallel(λQ.Q(boat2)•sank, P•A)

We can, by reasoning from the context, instantiate P to ‘λQ.Q(boat1)’, the only plausible candidate for a parallel item to B. The variable A must be instantiated to something distinct from ‘sank’ to meet the pragmatic condition of distinctness for ‘assert’. (Note that the identification of P and B would lead to a contradiction: thus Krifka’s statement of the pragmatic requirement on ‘assert’, as mentioned earlier, must be extended to include this case).

Thus an analogous chain of reasoning to the first case would lead us to deduce that something with similar relevant properties to 'sank' happened to the first boat: some other type of disaster, perhaps.

Thus the use of focus can lead to changes in the context. But can context influence the interpretation of focus? For this to be possible, focus must at least sometimes not be marked explicitly or unambiguously, for otherwise there would be little scope for contextual effects. In fact, it is clearly the case that intonational focus does not have to be explicitly marked: in written text, it seldom is, and even in spoken utterances there is no necessity for this to be the case. Particularly striking examples of this phenomenon are afforded by jokes which exploit the indeterminacy of focus.

111 a A. How many people work in this place? B: About half!
b A: Why do birds fly south in winter?
c B: Because it's easier than walking!

If focus were linguistically required to be overtly and unambiguously marked, such jokes would not be possible: it seems to be the case that in the absence of intonational clues the listener goes for the default VP focus, rather than narrow focus on the verb. Processing the second sentence reveals that this was the wrong interpretation and the sentence is reinterpreted. Clearly, if focus was always completely unambiguously marked, this reinterpretation, and the joke - such as it is - would not be possible.

We can illustrate the constraints that context places on intonational focus using the focus sensitive word 'too'. A provisional analysis of 'too' in the current framework would interpret it using an equivalence like:

too(F,S) $\Leftrightarrow$ S
if
Shared(F) = S,
& context(C),
& Shared(A) = C
& A$\approx$F.

Appropriate use of 'too' requires the utterance to share some content with an antecedent, but to contain something distinct as well.

Now consider an example in which we will assume, perhaps somewhat artificially, that focus is not specified at QLF. Assume that the context results from processing the sentence 'Joe sneezed' and that we are now processing 'He laughed too', having resolved the pronoun to 'Joe'.

too(F,S) $\Leftrightarrow$ S

if
$\quad$ Shared(F) = S, % S=laugh(joe),
$\quad$ & context(C), % C = sneeze(joe)
$\quad$ & Shared(A) = C % Shared = λQ.Q(joe); A = sneeze
$\quad$ & A$\approx$F. % sneeze$\approx$laugh

Focus is, by hypothesis, unspecified. At the stage where we have the equation 'Shared(F) = laugh(joe)' there are at least two potential solutions: Shared=laugh, F=joe; or Shared=λQ.Q(joe), F=laugh. However, the constraints provided by the context rule out the first of these, leaving the remaining possibility that we have to interpret focus as being on 'laugh'. This is consistent with the other conditions in the equivalence, and is intuitively correct: if focus is to be explicitly marked in this sentence, with this interpretation, that is where it must fall:

112 He LAUGHED too vs. *HE laughed too

Nevertheless, the context restricts the interpretation sufficiently that we arrive at the intended meaning without focus needing to be explicitly marked.

The combination of unification and abduction allows us to add new information to the context even when the context would be sufficiently well specified for normal purposes. Consider a sequence like:

113 Joe went to McDonalds. BILL had a bad meal too.

We have a (simplified) QLF 'too(bill,have(bill,a-bad-meal))'. In attempting to satisfy the conditions for the Too-focus equivalence we will find in the context only candidate sentences which will not satisfy the equations, in particular the equation 'Shared(A) = C'. However, we will be able to solve the equation 'Shared(F) = S', since we know the value of S and we know the value of F. This will give us 'λx.have(x,a-bad-meal)' as the value of Shared. In this context, we can now attempt to satisfy 'A$\approx$bill', and this should instantiate 'A=joe', since 'joe' will be the only salient individual in the context satisfying this property. Now we can reconstruct C from the equation 'Shared(A) = C', giving, after beta-reduction, 'have(joe,a-bad-meal)'. Adding this assumption to the context gives us exactly the piece of extra information that the speaker was intending to convey.

too(bill,have(bill,a-bad-meal)) $\Leftrightarrow$ have(bill,a-bad-meal)
if
$\quad$ Shared(F) = S, % Shared=λx.have(x,a-bad-meal),F=bill
$\quad$ & context(C), % C=[λx.have(x,a-bad-meal)(joe)
$\quad$ & Shared(A) = C % A=joe
$\quad$ & A$\approx$F. % joe$\approx$bill

While it is clearly not impossible for the other theories of focus discussed earlier to accommodate these interactions between focus and context, it seems to me that the higher order unification account, when supplemented with the conditional equivalence and associated reasoning mechanisms, does so in a simple and intuitively compelling way. (Some further interactions between focus, ellipsis, anaphoric reference, and context are explored in Pulman 1994).

3.3 Focus for Speech Synthesis

One of the very substantial advantages of the approach to focus that we have been pursuing is that it is completely reversible. Of course, any abstractly stated linguistic theory is in principle capable of supporting procedural descriptions of language processing in either the analysis or the synthesis direction. But in practice, there is often a large gap between the abstract statement of the theory and any actual computational implementation. One of the attractive properties of higher order unification as a logical tool is that it has quite a direct computational interpretation, and that interpretation is - even as an algorithm - independent of the direction of processing.

For our illustrations, we have been taking the point of view of a system for the analysis of utterances in some context. However, exactly the same mechanisms can be used in the reverse direction, for the synthesis of sentences with focus properties appropriate to the context in which they are uttered. In the case where focus can be expressed by pitch accent, there is an obvious practical application in the area of speech synthesis. (This has not escaped the attention of other researchers in this area, of course: see Prevost and Steedman 1993 for an application of Steedman's theory to synthesis). It is instructive to look at how a higher order equational analysis could be used to construct such a system.

Assume we have some interactive speech understanding application in which a user is interacting with a system in order to operate some machinery. Envisage a scenario like the following:

114 User to system: Do I put the card into the green slot?

The system analyses this as:

115 $\exists$xy. put(user,x,y) & card(x) & slot(y) & green(y)

Assume that the correct operation at this point is actually that you put a disc into a red slot:

116 $\exists$xy. put(user,x,y) & disc(x) & slot(y) & red(y)

Now the system is in a position to infer the QLF of various appropriate responses, by applying the conditional equivalences in the right to left direction. For example, the system can infer that the relevant components of the correct action should be given narrow focus (with contrastive stress), by using the equivalence for positive focus sentences, 'assert', where the equations are resolved as follows (where C is instantiated to example 115, and S to example 116).

assert(disc•red,S) $\Leftrightarrow$ S
    if
        B(F1•F2) = S,
        % [λPQ. ∃xy.put(user,x,y) & P(x) & slot(y) & Q(y)](disc,red)
        & context(C),
        & P(A1•A2) = C,
        % [λPQ. ∃xy.put(user,x,y) & P(x) & slot(y) & Q(y)](card,green)
        & parallel(B•disc•red, P•card•green).

Here the user's utterance is the context, and the right hand side of the equivalence is the correct information that needs to be expressed in a contextually appropriate way.

Once we have the QLF representing the focus structure of the proposition it is a relatively simple matter to use a grammar like that in our earlier fragment to generate the sentence:

117 You put the DISC into the RED slot

Other variants on these themes are possible, and it is not too difficult to extend this scenario to allow for uses of focussing particles like 'even' and 'only'.

If the amount of higher order inference needed to work out an adequate response becomes too great, there may be some practical difficulties about producing an efficient implementation along these lines, at least in the short term. I nevertheless think that an approach using higher order unification is currently the best candidate for a general, principled, declarative, and computationally grounded solution to the problem of the production and interpretation of sentences with intonational or other focus appropriate to their contexts.

4 Bibliography

Acknowledgements

The present paper is a much revised and expanded version of one given to the Third Nordic Conference on Text Comprehension in Man and Machine, April 1992, Linköping University, which appeared as Pulman 1993. Versions of the paper were also given during 1992 to audiences at the Universities of Cambridge, Essex and Sussex; to the Artificial Intelligence Center, SRI International, Menlo Park, California; to a conference on Computational Semantics organised by Stuttgart University in Freudenstadt, Germany; and to a meeting of the European Science Foundation on Logic and Language, in Autrans, France. I thank members of the audience on all these occasions for their helpful suggestions. Thanks are also due to Mark Steedman for some points of clarification (I doubt he will agree with my conclusions) and to two anonymous referees for their astute comments on an earlier version.

Preparation of this version of the paper was partially supported by the FraCaS project, DG XIIIb, Luxembourg, and the UK Defence Research Agency under contract CSM3/002.

Bibliography

Alshawi, H. (ed) 1992, The Core Language Engine, Cambridge, Mass: MIT Press.

Alshawi, H. and Crouch, R., 1992, *Monotonic Semantic Interpretation*, Proceedings of the 30th Annual Meeting, Association for Computational Linguistics, pp 32-39.

Alshawi, H. and Pulman, S. G., 1992, *Ellipsis, Comparatives and Generation*, in H. Alshawi (ed) The Core Language Engine, Boston: MIT Press, pp 251-275.

Carter, D., 1987, *Interpreting Anaphors in Natural Language Texts*, Ellis Horwood: Chichester.

Chomsky, N. 1970, Deep Structure, Surface Structure, and Semantic Interpretation, in 'Semantics: an Interdisciplinary Reader', ed. D.D. Steinberg and L.A. Jacobovits, Cambridge, CUP, 1971, pp 183-216.

Dalrymple, M., Shieber, S. and Pereira, F., 1991, *Ellipsis and Higher Order Unification*, Linguistics and Philosophy, 14: 399-452.

Davidson, D. 1967, The Logical Form of Action Sentences, in The Logic of Decision and Action, ed. N. Rescher, Pittsburgh: University of Pittsburgh Press, pp 81-95.

Delin, J., 1989, Accounting for Cleft Constructions in Discourse: a Multi-Layered Approach. Edinburgh University Human Communication Research Centre Research Paper HCRC/RP-5.

Dretske, F. 1979 Referring to Events, in P. French, T. Uehling, and H. Wettstein, eds., Contemporary Perspectives in the Philosophy of Language, Minneapolis: University of Minnesota Press, 1979: pp 369-378.

Gawron, J. M. 1992 *Focus and Ellipsis in Comparatives and Superlatives: a Case Study*, in Proceedings of the Second Conference on Semantics and Linguistic Theory, Working Papers in Linguistics No 40, Ohio State University, ed. C.

Barker and D. Dowty, pp 79-98.

Gawron, J. M. 1995 Comparatives, Superlatives, and Resolution, Linguistics and Philosophy, 18: ??? -380.

Gazdar, G., Klein, E., Pullum, G. and Sag, I. 1985 Generalised Phrase Structure Grammar, Oxford: Basil Blackwell.

Hajičová, E. 1984, *Topic and Focus*, in P. Sgall (ed), *Contributions to Functional Syntax, Semantics, and Language Comprehension*, Czechoslovak Academy of Sciences: Prague, pp 189-202.

Horn, L. R. 1969, A presuppositional analysis of ‘only’ and ‘even’, in Papers from the 5th Regional Meeting of the Chicago Linguistics Society, ed. R. Binnick, A. Davidson,G. Green and J. Morgan: 98-107.

Huet, G. 1975, *A Unification Algorithm for Typed λ-Calculus*, Theoretical Computer Science, 1, pp 27-57.

Jackendoff, R. 1972, *Semantic Interpretation in Generative Grammar*, MIT Press, Boston.

Joshi, A. 1990, Phrase Structure and Intonational Phrases, in G. Altmann (ed) Cognitive Models of Speech Processing, Cambridge, Mass: MIT Press, pp 457-482.

Kratzer, A. 1991, The Representation of Focus, in A. von Stechow and D. Wunderlich (eds) Semantics: An International Handbook of Contemporary Research, Berlin: de Gruyther.

Krifka M. 1990 Four thousand ships passed through the lock, Linguistics and Philosophy, 13, 487-520.

Krifka, M. 1991, A Compositional Semantics for Multiple Focus Constructions, Linguistische Berischte Sonderheft 4, 17-54.

Krifka, M. 1992, A Framework for Focus-Sensitive Quantification, in Proceedings of the Second Conference on Semantics and Linguistic Theory, Working Papers in Linguistics No 40, Ohio State University, ed. C. Barker and D. Dowty, pp 213-236.

Krifka, M. 1993, Focus and Presuppositions in Dynamic Interpretation, Journal of Semantics, 10, 269-300.

Lewis, D. 1975 Adverbs of Quantification, in E. Keenan (e) Formal Semantics of Natural Language, Cambridge: Cambridge University Press.

Lyons, D. and Hirst, G. 1990, *A Compositional Semantics for Focusing Subjuncts*, Proceedings of the 28th Annual Meeting, Association for Computational Linguistics, pp 54-61.

Miller, D. and Nadathur, G. 1986, *Some uses of higher order unification in computational linguistics*, in Proceedings of the 24th Annual Meeting, Association for Computational Linguistics, pp 247-255.

Partee, B. 1991, Topic, Focus, and Quantification, in Proceedings of the 1st Semantics and Linguistic Theory Conference, ed. S. Moore and A. Z. Wyner, Cornell University Working Papers in Linguistics 10, pp 159-187.

Prevost, S. and Steedman, M. 1993 Generating Contextually Appropriate Intonation, on Proceedings of the 6th Conference of the European Chapter of the ACL, Utrecht: ACL, pp 332-340.

Prüst, H. 1992 *On Discourse Structure, VP Anaphora, and Gapping*, University of Amsterdam Ph.D. Dissertation.

Pulman, S. G. 1989 'Events and Verb Phrase Modifiers', in Recent Developments and Applications of Natural Language Processing, ed. J. Peckham, London: Kogan Page Ltd, and Maryland: GP Publishing, pp 179-191.

Pulman, S. G. 1991, *Comparatives and Ellipsis*, Proceedings of the 5th European Meeting of the Association for Computational Linguistics, Berlin, 1991, pp 1-6.

Pulman, S. G. 1993, *Higher Order Unification and the Semantics of Focus* in Papers from the Third Nordic Conference on Text Comprehension in Man and Machine, April 1992, Linköping University, 113-127.

S. G. Pulman, 1994, *A Computational Theory of Context Dependence*, in Proceedings: International Workshop on Computational Semantics, ed. H. Bunt, R. Muskens, and G. Rentier, Institute for Language Technology, Tilburg University, The Netherlands, 161-170.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. 1985 A Comprehensive Grammar of the English Language, Harlow: Longman Group Limited.

Rayner, M. 1993, *Abductive Equivalential Translation and its application to Natural Language Database Inferencing,* Stockholm Univ. Ph.D. thesis, and SRI Cambridge Technical Report.

Rayner, M. and Alshawi, H. 1992 *Deriving Database Queries from Logical Forms by Abductive Definition Expansion*, in Proc. 3rd Conference on Applied Natural Langauge Processing, Trento: ACL pp 1-8.

Rooth, M. 1985, Association with Focus, Ph.D. Thesis, University of Massachusetts, Amherst.

Rooth, M, 1992, A Theory of Focus Interpretation, Natural Language Semantics, 1, pp 75-116.

Schubert, L. and Pelletier, J. 1989, Generically Speaking, in G. Chierchia, B. Partee, and R. Turner, (eds) Properties, Types, and Meaning, II, Dordrecht:

Kluwer Academic Publishers, pp 193-268.

Selkirk, E. 1984, *Phonology and Syntax*, MIT Press: Boston.

Sgall, P. 1984, *Syntax, Meaning, and Reference*, in P. Sgall (ed), *Contributions to Functional Syntax, Semantics, and Language Comprehension*, Czechoslovak Academy of Sciences: Prague, pp 39-47.

Sidner, C. 1983, Focussing in the Comprehension of Definite Anaphora, in 'Computational Models of Discourse', ed. M. Brady and R. Berwick, Cambridge, Mass: MIT Press, pp 276-330.

Steedman, M. 1987 Combinatory Grammars and Parasitic Gaps, Natural Language and Linguistic Theory, 5, 403-439.

Steedman, M. 1990 Structure and Intonation in Spoken Language Understanding, in Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics, Pittsburgh: ACL, pp 9-16.

Steedman, M. 1991 Structure and Intonation, Language 68, 260-296.

Steedman, M. 1993 The Grammar of Intonation and Focus, in Proc. 9th Amsterdam Colloquium, University of Amsterdam ILLC, pp 17-33.

Stump, G. 1981 The interpretation of frequency adjectives, Linguistics and Philosophy 4, 221-257.

de Swart, H. 1991 Adverbs of Quantification: A Generalised Quantifier Approach, Univ. of Groningen Ph.D. Thesis