

SIGRID BECK AND HOTZE RULLMANN

A FLEXIBLE APPROACH TO EXHAUSTIVITY
IN QUESTIONS*

A semantics for interrogatives is presented which is based on Karttunen's theory, but in a flexible manner incorporates both weak and strong exhaustivity. The paper starts out by considering degree questions, which often require an answer picking out the maximal degree from a certain set. However, in some cases, depending on the semantic properties of the question predicate, reference to the minimal degree is required, or neither specifying the maximum nor the minimum is sufficient. What is needed is an operation which defines the maximally informative answer on the basis of the Karttunen question denotation. Shifting attention to non-degree questions, two notions of answerhood are adopted from work by Heim. The first of these is weakly exhaustive and the second strongly exhaustive. The second notion of answerhood is proven to be equivalent to Groenendijk and Stokhof's interrogative semantics. On the basis of a wide range of empirical data showing that questions often are not interpreted exhaustively, it is argued that a fairly rich system of semantic objects associated with questions is needed to account for the various ways in which questions contribute to the semantics and pragmatics of the utterances in which they appear.

1. INTRODUCTION

In this paper we propose a modification and extension of Karttunen's (1977) semantics for interrogatives which incorporates a flexible approach to the property of questions called exhaustivity.¹ Karttunen's original proposal was criticized (in particular by Groenendijk and Stokhof 1982, 1984) for failing to account for what has been termed strong exhaustivity. This is a property of embedded questions that licenses inferences like (1), which Groenendijk and Stokhof claim are valid:

- (1) John knows who was at the party.
Mary was not at the party.
∴ John knows that Mary was not at the party.

* While we wrote this paper, the first author was supported by a DFG grant to the SFB 340 and the second author by a grant from the Dutch Organization for Scientific Research (NWO), as part of the PIONIER-project 'Reflections of Logical Patterns in Language Structure and Language Use' at the University of Groningen, and by an Izaak Walton Killam Memorial Postdoctoral Fellowship at the University of Alberta. We are grateful to Irene Heim, Jack Hoeksema, Paul Portner, Arnim von Stechow, Wolfgang Sternefeld, the audience at the Tenth Amsterdam Colloquium, and two anonymous reviewers for comments and discussion.

¹ Throughout this paper we use the terms 'question' and 'interrogative' interchangeably to refer to syntactic objects in natural language. When we want to refer to the corresponding semantic objects we will use terms like 'denotation' or 'intension'.

They propose a different semantic analysis of interrogatives, which accounts for (1) by virtue of the basic question interpretation. Their point is accepted in Rullmann (1995), who proceeds to make a proposal that accounts for strong exhaustivity within a Karttunen system. His means of implementing strong exhaustivity is a maximality operator. Rullmann's motivation for the assumption of a maximality operator is not only exhaustivity, but also an effect in degree questions that we refer to as the 'maximality effect'. Degree questions like (2) below seem to require an answer that is in some sense maximal:

- (2) How many books did John read?

That is, someone who asks (2) will only be satisfied with an answer specifying the largest number of books that John read. This effect is captured in Rullmann (1995) by making maximality part of the basic question denotation by means of a maximality operator. While we agree that maximality and exhaustivity should be viewed as one and the same phenomenon, we suggest a different way of accounting for them.

We will defend a Hamblin/Karttunen-style semantics for questions, in which the basic denotation of a question is a set of propositions which intuitively constitute its possible answers. However, we will also incorporate Groenendijk and Stokhof's insights about what information questions introduce in embedded constructions by adopting a proposal by Heim (1994). She accounts for properties like strong and weak exhaustivity by defining two semantic notions of answerhood (which she calls *answer1* and *answer2*), which create propositions from Hamblin/Karttunen question intensions. We show that this also captures the maximality effect in degree questions.

Heim's two notions of answerhood provide us with a fairly rich system of semantic objects definable in terms of the basic question denotation. We argue that this rich system is needed in the analysis of interrogative constructions in natural language. Our proposal is superior to Groenendijk and Stokhof's and Rullmann's in that their interrogative semantics does not make available all semantic objects that the analysis of interrogatives requires, as there appears to be considerable variation in what an interrogative contributes to a construction semantically.

We will take as our starting point Rullmann's (1995) proposal incorporating strong exhaustivity with the help of a maximality operator (section 2). Although part of the motivation for that operator (besides strong exhaustivity) comes from the maximality effect in degree questions, we will see that once a broader range of degree questions is considered, the assumption of a maximality operator is actually problematic (section 3). On the

EXHAUSTIVITY IN QUESTIONS

other hand, all types of degree questions can receive a satisfactory treatment in terms of Heim's notions of answerhood. In section 4, we introduce Heim's first notion of answerhood (answer1) and show in some detail how it can account for the data which are problematic for the maximality account. In section 5, we turn to exhaustivity. Heim's approach reanalyzes exhaustivity not as a property of the question, but as a property of the notion of a true, complete answer to the question. We prove that her second notion of answerhood (answer2) is equivalent to Groenendijk and Stokhof's interrogative semantics, once certain complications dealing with so-called *de dicto* and *de re* readings of *which*-phrases are dealt with (section 6). This means that we can account for exhaustivity and maximality effects even with a basic Hamblin/Karttunen semantics as our starting point. However, the differences between our proposal and Groenendijk and Stokhof's semantics are not merely a matter of implementation. Section 7 is devoted to the differences between the two proposals and their potential empirical implications. This includes discussion of well-known phenomena (like different kinds of question-embedding verbs and mention-some interpretations) as well as some new data (in particular a type of degree question that involves *at least* and *at most*). Section 8 discusses two potential ways of implementing our suggestions in the syntax/semantics interface. We conclude that our proposal ought to be considered as an alternative, flexible approach to exhaustivity and maximality.

2. AN APPROACH BASED ON A MAXIMALITY OPERATOR

2.1. Karttunen (1977)

Before discussing Rullmann's (1995) implementation of maximality, we will briefly sketch Karttunen's semantics for questions on which his account is based. For a full exposition the reader is referred to Karttunen (1977). In Karttunen's semantics the denotation of a question is a set of propositions, namely the set of all those propositions that are true answers to the question. If for instance Mary, Sue, and Jane were at the party and no one else was, then the denotation of the question (3a) will be the set of propositions given in (3b):

- (3) a. Who was at the party?
b. {Mary was at the party, Sue was at the party, Jane was at the party}
c. $\lambda p \exists x [\text{person}(w)(x) \ \& \ p(w) \ \& \ p = \lambda w' [x \text{ was at the party in } w']]$

More generally, in a world w (3a) denotes the set of propositions (3c), which

we can think of informally as the set of propositions of the form 'x was at the party' which are true in w , where x is a person (though in reality of course propositions can't be said to have a 'form', because they are sets of possible worlds).²

We will use the term 'Karttunen denotation' to refer to the extension of a question (i.e., to an object of type $\langle\langle s, t \rangle, t \rangle$). When we are talking about Karttunen intensions (i.e., objects of type $\langle s, \langle\langle s, t \rangle, t \rangle \rangle$), this should be clear from the context.

2.2. Degree Questions and Maximality

In chapter 3 of his dissertation, Rullmann (1995) notes that degree questions like (4a) and (4b) require an answer that is in some sense maximal:

- (4) a. How many books did John read?
 b. How high can John jump?
 c. Jill knows how high John can jump.

Someone who utters (4a) wants to know the maximal number n such that John read n books. Similarly, (4b) asks for the maximal (degree of) height d such that John can jump d -high. The embedded case (4c) is parallel: Jill has to be aware of the maximal height John can jump. Note that if John read five books and not more than five books, then the only possible true answer to (4a) will be 'five', even though the proposition that John read four books is literally speaking true in that situation. We will call this effect maximality.

Rullmann's idea is that (4a) and (4b) really mean something like (5a,b):

- (5) a. Which number n is such that n is the greatest number of books that John read?
 b. Which degree d is the greatest degree such that John can jump d -high?

Quasi-formally, (5a) and (5b) can be represented as in (6a) and (6b), where 'max' is an operator that picks out the maximal degree (or number) from a set of degrees and '?' is a question-operator whose semantics is spelled out below:

² We assume an extensional logical language with explicit variables standing for possible worlds. For perspicuity we adopt the convention that the world variable is always written in parentheses as the first argument of a predicate. When free, the variable w refers to the actual world.

EXHAUSTIVITY IN QUESTIONS

- (6) a. ?n: $n = \max(\lambda n'[\text{John read } n' \text{ books}])$
 b. ?d: $d = \max(\lambda d'[\text{John can jump } d' \text{-high}])$

This basic idea can be implemented in a Karttunen-style semantics of questions, where the denotation of a question is the set of propositions that are true answers to the question:

- (7) a. $\lambda p \exists n [p(w) \ \& \ p = \lambda w' [n = \max(\lambda n' [\text{John read } n' \text{ books in } w'])]]$
 b. $\lambda p \exists d [p(w) \ \& \ p = \lambda w' [d = \max(\lambda d' [\text{John can jump } d' \text{-high in } w'])]]$

(7a,b) are basically the Karttunen denotations of the paraphrases in (5a,b). Note that these formulas will always denote a singleton set (or the empty set), because there is at most one maximal degree or number. The single element is the maximum answer, so Rullmann's theory accounts for the maximality effect. One immediate objection that might be raised against this analysis is that it builds maximality into the semantics of questions with no role to play for pragmatics, thereby separating this phenomenon from that of scalar implicatures in declarative sentences. We will leave this issue aside for the moment. The theory we eventually argue for is one which does leave more room for pragmatics, especially in the analysis of unembedded questions (see section 8.3).

In the rest of the paper we will mostly give examples of degree questions involving numbers (*how many*) rather than degrees, for reasons of simplicity.

2.3. *Individual Questions and Exhaustivity*

Rullmann shows that his analysis can be extended to questions involving individuals rather than degrees or numbers, if we adopt a Link-style analysis in which the domain of discourse contains not only atomic individuals, but also their mereological sums, or groups. Maximality should then be interpreted with respect to the 'part-of' relation on groups. A question like (8) can be analyzed as asking for the maximal group of individuals such that this group was at the party:

- (8) a. Who was at the party (last night)?
 b. Which x is such that x is the largest group that was at the party?
 c. ? x : $x = \max(\lambda x' [x' \text{ was at the party}])$
 d. $\lambda p \exists x [p(w) \ \& \ \text{person}(w)(x) \ \& \ p = \lambda w' [x = \max(\lambda x' [x' \text{ was at the party in } w'])]]$

Following a suggestion by Jacobson (1995), Rullmann shows that maxi-

maximality can be used to account for the property of strong exhaustivity argued for by Groenendijk and Stokhof (1982, 1984). Groenendijk and Stokhof distinguish two kinds of exhaustivity in questions: weak and strong exhaustivity.³ Weak exhaustivity is the property which licenses inferences of the following form:

- (9) John knows who was at the party.
 Mary was at the party.
 ∴ John knows that Mary was at the party.

Strong exhaustivity is the property of questions which makes it possible to draw inferences of the following type (in addition to ones like (9)):

- (10) John knows who was at the party.
 Mary was not at the party.
 ∴ John knows that Mary was not at the party.

Groenendijk and Stokhof (1982) propose a semantics for questions which is an alternative to Karttunen's theory, but which accounts for both weak and strong exhaustivity (unlike Karttunen's analysis, which captures only weak exhaustivity).

Rullmann (1995) argues that by introducing maximality into the Karttunen semantics for questions we get an analysis that, like Groenendijk and Stokhof's theory, accounts for both weak and strong exhaustivity. As noted above, maximality guarantees that a question will always denote a singleton set of propositions. Because there is a one-to-one relation between singleton sets and their elements, we may therefore as well identify the denotation of a question with the proposition that is the unique member of this set (at least for the class of question-embedding verbs which can take *that*-clause complements and which Groenendijk and Stokhof call extensional, such as *know*). This means that the denotation of (8a) will be the proposition in (11), rather than the singleton set containing it:

- (11) $\uparrow p \exists x [p(w) \ \& \ \text{person}(w)(x) \ \& \ p = \lambda w' [x = \max(\lambda x' [x' \text{ was at the party in } w'])]]$

Now suppose that in the actual world *w*, Mary, Sue, and Jane were at the party and no one else was. Then the proposition denoted by (11) will be:

³ In fact, Groenendijk and Stokhof don't use the terms 'strong' and 'weak' exhaustivity, but discuss various 'degrees' of exhaustivity, also distinguishing a third, intermediate degree. According to Berman (1991) the terms 'strong' and 'weak' exhaustivity were first used in a paper by Bäuerle and Zimmermann.

EXHAUSTIVITY IN QUESTIONS

- (12) $\lambda w'[\text{Mary} + \text{Sue} + \text{Jane} = \max(\lambda x'[x' \text{ was at the party in } w'])]$

This proposition contains all and only those worlds in which the party-goers are Mary, Sue, and Jane, and no one else. Now if John stands in the 'know'-relation to this proposition, this will imply that for every x such that x is a member of {Mary, Sue, Jane}, John knows that x was at the party, and that for every x such that x is not a member of {Mary, Sue, Jane}, John knows that x was not at the party. (Crucially, just like Groenendijk and Stokhof we assume that knowing p entails knowing every proposition entailed by p .) In other words, maximality accounts for both weak and strong exhaustivity. Thus, by adding maximality to Karttunen's theory of questions, we end up with a theory that – though not formally equivalent to it – is able to account for the intuitions that motivate Groenendijk and Stokhof's theory.

Note that strong exhaustivity plays a role in degree questions in just the same way that it does in individual questions. If John knows how many books Bill read, and in fact Bill read five books and not more than five, then by strong exhaustivity, for any $n > 5$, John knows that Bill did not read n books. Rullmann's (1995) analysis of degree questions accounts for this implication in the same way that it does for individual questions.

In this paper, we build on and substantially modify Rullmann's proposal in order to deal with a class of degree questions in which minimality rather than maximality seems to be called for. We show that if we interpret maximality as maximal informativeness and apply it at the level of propositions rather than degrees or individuals, we can account for this class of examples, while at the same time preserving the intuition that maximality and exhaustivity are two sides of the same coin. In the second half of the paper, we then show that by adopting this perspective we get a theory which allows for a more flexible approach to weak and strong exhaustivity than the one defended by Groenendijk and Stokhof. In this respect we largely follow Heim (1994). In support of this approach we will discuss data (in part derived from the existing literature) which show that degree and non-degree questions are not uniformly interpreted exhaustively. Finally, we suggest two ways in which non-exhaustive and exhaustive interpretations of questions may be related in a theory either lexically (by means of meaning postulates or lexical decomposition) or by type shifting.

3. PROBLEMS WITH THE MAXIMALITY OPERATOR

3.1. *Degree Questions Requiring a Minimal Answer*

Consider a question like (13).

- (13) How many eggs are sufficient (to bake this cake)?

Intuitively, if you ask (13), you want to know the smallest number n such that n eggs would be sufficient to bake the cake. The interpretation we would get for (13) in Rullmann's (1995) analysis, however, is given in (14).

- (14) $?n: n = \max(\lambda n'[n' \text{ eggs are sufficient to bake this cake}])$

This is not a satisfactory interpretation for (13), for two (related) reasons. First, the maximum is likely to be undefined in this case: Suppose that in fact three eggs are sufficient to bake the cake, but fewer than three eggs are not. Then four eggs are also sufficient, and so are five, six, etc. So there is no largest number of eggs that would be sufficient. Let us ignore this problem for a moment, though; maybe the set of numbers is contextually restricted in some way, so that a largest element is defined. Even then, we do not end up with the desired interpretation for the question, because this gives us the largest number of eggs that would be sufficient, while we intuitively want the smallest such number, namely three. If we formalize (13) in a way analogous to (4a) in section 2.2, a more appropriate solution would be (15):

- (15) $?n: n = \min(\lambda n'[n' \text{ eggs are sufficient to bake this cake}])$

There are a few other predicates that intuitively behave in the same way as *be sufficient*:

- (16) a. Mit wieviel Geld kann ein Professor auskommen?
with how-much money can a professor make-do
'On how much money can a professor live?'
- b. Wie weit zu schwimmen ist ausreichend?
how far to swim is sufficient
'How far is it sufficient to swim?'
- c. Wieviel Arsen kann einen Menschen umbringen?
how-much arsenic can a man kill
'How much arsenic is enough to kill somebody?'
- d. How big a difference (in light intensity) is perceptible?

EXHAUSTIVITY IN QUESTIONS

In all these examples an appropriate answer would name a minimum (the minimal amount of money on which a professor can live, the minimal distance it suffices to swim, etc.), rather than a maximum. Why should that be the case?

The 'minimum' interpretation crucially depends on what we will call the question predicate. We will somewhat informally use this term to refer to what is the argument of the max-operator in formulas like (14). In the degree questions in section 2 that required maximal answers, we always had question predicates that allowed inferences from larger degrees to smaller ones. So for instance in (17), the question predicate (17b) allows inferences from a number n to numbers m smaller than n ; thus, if John has read five books, then he has also read four books, three books, etc.

- (17) a. How many books did John read?
b. $\lambda n'[\text{John read } n' \text{ books in } w']$

So in (17) the question predicate has the following property:

- (18) A predicate P is downward scalar iff $\forall n, m[[P(n) \ \& \ m \leq n] \rightarrow P(m)]$

In the minimality inducing examples (13) and (16), on the other hand, the question predicate has the reverse property:

- (19) A predicate P is upward scalar iff $\forall n, m[[P(n) \ \& \ n \leq m] \rightarrow P(m)]$

For instance, if three eggs are sufficient, then four eggs, five eggs, etc. will also be sufficient. The upward scalar predicates seem to be considerably rarer than the downward scalar ones. We will discuss how these scalar properties come about in section 4.

We think that the difference between the maximality inducing examples and the minimality inducing ones boils down to informativeness. In case the question predicate allows inferences from a large number to smaller ones, the most informative answer to the question will be to name the maximum, since this implies all other true answers. In the minimality case, it is most informative to give the minimum answer, because here the minimum implies all other true answers. Therefore, we believe that it is misguided to give the maximum (or, for that matter, the minimum) any special status. We conclude that we should have neither a maximum- nor a minimum-operator in the semantics of degree questions. Note that we do not get an ambiguity; what type of answer is required seems fixed for a given predicate. The fact that we choose the maximum answer in the case of downward scalar predicates should follow from general principles.

The same principles should account for the fact that upward scalar predicates require a minimum answer.

So far, our remarks on informativeness have been completely informal. Before we turn to a proper formalization of our idea, we would like to discuss another type of question predicate that behaves in yet another way.

3.2. *Degree Questions with Non-Scalar Predicates*

Consider (20):

- (20) a. How many people can play this game?
- b. How many people can form a soccer team?
- c. How many processors can Windows NT support?

A complete answer to (20a) could be, for instance, 'between 4 and 6'. Or a certain game may be played with any even number of players. This is, in effect, a complete list of true answers, or to put it differently, their conjunction. Similarly for the other examples. In (20c), for instance, the true answers (we are assured by Thilo Goetz, p.c.) are 1, 2, and 4. The question predicates in (20) are predicates that do not allow inferences either from larger degrees to smaller ones or the other way around. Hence, the question predicates in (20) are neither downward scalar nor upward scalar. We will refer to them as non-scalar predicates. Since in these cases naming one true answer does not allow any inferences, the only fully informative answer is the conjunction of all true answers. So this is a case where neither a maximum nor a minimum operator would get us anywhere. Resorting to informativeness, however, is still a natural thing to do.

In (21), there are a few more questions in which intuitively the question predicate does not allow for any inferences:

- (21) a. How many courses are you allowed to take per semester?
- b. How high can a helicopter fly?

You might be allowed to take either 4 or else between 6 and 8 courses per semester, depending on the program you are in. And a helicopter might have a minimum as well as a maximum of altitude. While we will derive the intuitive inferential behavior of (20), the data in (21) will not be analyzed in any detail. They are provided to strengthen the point that intuitively, informativeness is at stake.

EXHAUSTIVITY IN QUESTIONS

4. MAXIMAL INFORMATIVENESS OF ANSWERS

4.1. *Answer1*

In this section we give a more formal implementation of the view that informativeness is the crucial notion in describing the types of answers appropriate for degree questions. Our strategy will be to incorporate informativeness not into the basic semantics of the question itself, but into the definition of answerhood to a question.

Just like Karttunen, we take the basic denotation of a question to be a set of propositions, which intuitively represent (partial) answers to the question. However, we slightly depart from Karttunen in that we do not require that the propositions in this set be true. Thus, we follow Hamblin (1973) and assume that the basic denotation of a question includes both true and false propositions. (22a), for example, will have (22b) as its basic denotation instead of (22c):

- (22) a. Who was at the party?
b. $\lambda p \exists x [\text{person}(w)(x) \ \& \ p = \lambda w' [x \text{ was at the party in } w']]$
c. $\lambda p \exists x [\text{person}(w)(x) \ \& \ p(w) \ \& \ p = \lambda w' [x \text{ was at the party in } w']]$

The reason for our departure from Karttunen is that some question-embedding verbs – e.g., *agree* – are sensitive to possible answers rather than to true answers, as will be discussed in section 6.2. Although representations like (22b) incorporate one important aspect of Hamblin's account, in other respects they are still more in the spirit of Karttunen's work. (Later on in this paper, we will actually introduce a further modification which takes us closer to Hamblin's original semantics.) We will therefore refer to (22b) as the Hamblin/Karttunen denotation of the question, and to the corresponding intension as the Hamblin/Karttunen intension.

Starting from the Hamblin/Karttunen question denotation, we define the concept of a maximally informative answer. As it turns out, this notion has already been formalized by Heim (1994), who calls it *answer1*. Let Q be a Hamblin/Karttunen intension. Then the *answer1* to Q in a world w ($\text{answer1}(w)(Q)$) is defined as follows:

$$(23) \quad \text{answer1}(w)(Q) = \bigcap \{p: Q(w)(p) \ \& \ p(w)\}$$

This is the conjunction of all true propositions in the question extension. Later on, we will discuss a second concept of answerhood, *answer2*. Related notions of answerhood have been proposed by Lahiri (1991) and Dayal (1996).

To see how *answer1* works, and how it explains the scalarity effects

discussed above, we will in the remainder of this section discuss an example of each of our three types of question predicate (upward scalar, downward scalar, and non-scalar) in some detail. But first we provide a few additional background assumptions.

Throughout the paper, we assume a plural ontology in which we have group individuals as well as singular individuals. The domain of individuals is ordered by a part-of relation ' \preceq '. We have a predicate 'card' which assigns to a group the exact number of atomic subparts of that group. We will also assume standard pluralization of predicates with the *-operator (see Link (1983) for discussion and explicit definitions):

- (24) For any predicate P, *P includes all individuals x such that P(x) as well as all groups that can be formed from these individuals.

Moreover, for NPs like *five books*, and also *n books*, we will consistently assume an interpretation that amounts to an 'at least' reading, e.g:

- (25) *five books* $\rightsquigarrow \lambda P \lambda w \exists X [*book(w)(X) \ \& \ card(X) = 5 \ \& \ P(w)(X)]$

We might not make the intended semantics always this explicit in our informal statements; in what follows, when we talk about the proposition 'John read five books', we will mean that he read at least five books.

In this paper we will ignore the semantic difference between singular and plural *wh*-phrases. For a proper discussion of the effect of morphological number on *wh*-phrases see Dayal (1996), which is compatible with our discussion.

4.2. A Downward Scalar Predicate

Let's start with a 'standard' downward scalar predicate. The question in (26) has the Hamblin/Karttunen denotation in (27a,b).

- (26) How many books did John read?
- (27) a. $\lambda p \exists n [p = \lambda w' [John \text{ read } n \text{ books in } w']]$
 b. $\lambda p \exists n [p = \lambda w' \exists X [*book(w')(X) \ \& \ card(X) = n \ \& \ * \lambda y [read(w')(y)(John)](X)]$

Suppose that John actually read five books, and no more than five. Then, the set of propositions that are elements of (27b) and that are true are given in (28):

- (28) {John read five books, John read four books, John read three books, John read two books, John read one book}

EXHAUSTIVITY IN QUESTIONS

This is so since we have assumed what amounts to an 'at least' interpretation for the NP, and *read* is distributive: if you read a set of five books, you read each of them. Hence, if there is a group of five books that you read, there is also a group of four books that you read.

Answer1, apart from filtering out the false propositions, forms the intersection of all the propositions in the resulting set, that is, their conjunction. In the case we are looking at, however, the proposition that John read five books is a subset of the proposition that he read four books, and similarly for all the other propositions that are true. So the intersection of all these propositions is the same set of possible worlds as the proposition that John read five books. Answer1(w)((27b)) is thus identical to the maximum answer.

4.3. An Upward Scalar Predicate: Sufficient

We now turn to a case with an upward scalar predicate, example (29), which we showed to be problematic for Rullmann's account because a minimum answer is required:

- (29) How many eggs are sufficient?

Our predictions for this case depend on the analysis of the predicate *be sufficient*. In terms of our overall goals, the following is a slight digression, but it serves to make our own predictions explicit. We try to keep the discussion short, though.

We suggest that (30) means (31a) or equivalently (31b):

- (30) Four eggs are sufficient (to bake this cake).
- (31) a. It is not necessary (given the rules for your cake baking) that you have more than four eggs.
b. It is possible (given the rules for your cake baking) that you have only four eggs.

We will derive this semantics via the lexical meaning of *sufficient*. We will take as our guideline the paraphrase in (31b). We will assume that semantically the argument of *sufficient* is propositional in nature. *Sufficient* then contributes modal possibility as well as a meaning component amounting to *only*. Thus, we suggest the following semantics for *sufficient*:

- (32) $\llbracket \text{sufficient} \rrbracket^{\circ}(w)(p) = 1$ iff $\exists w': w \sim w' [\neg \exists q [q \in \llbracket p \rrbracket^f \ \& \ q(w') \ \& \ \llbracket p \rrbracket^{\circ} \nrightarrow q]]$

That is, we propose that *sufficient* is sensitive to the focus semantic value of its propositional argument as well as to its ordinary semantic value. It acts on this information much like *only*. We follow Rooth's (1985, 1992) theory of focus. With this semantics, we predict focus sensitivity of *sufficient*. We think that this is accurate: (33a) and (b) are truth-conditionally different.

- (33) a. It is sufficient to introduce Mary to JOHN.
 b. It is sufficient to introduce MARY to John.

Sentence (33a) means (34):

- (34) $\exists w': w' \sim w[\neg \exists q[q \in \llbracket \text{you introduced Mary to [John]}^f \rrbracket \& q(w') \& \lambda w''[\text{you introduced Mary to John in } w''] \nrightarrow q]]$

This can be simplified to (35):

- (35) $\exists w': w' \sim w[\neg \exists q[q \in \{\lambda w''[\text{you introduced Mary to } y]: y \text{ an individual}\} \& q(w') \& q \neq \lambda w''[\text{you introduced Mary to John in } w'']]]$

This means that there is a permissible state of affairs in which there is no focus alternative 'You introduced Mary to y' that is true other than 'You introduced Mary to John'. By the same reasoning, (33b) means that there is a permissible state of affairs in which there is no focus alternative 'You introduced z to John' that is true other than 'You introduced Mary to John'. Hence in (33a), you needn't introduce Mary to anyone else (but might have to introduce other people to John), while in (33b), you needn't introduce anyone other than Mary to John (but might have to introduce Mary to other people).

Our semantics predicts that (36a) implies (36b), given a focus on *four*:

- (36) a. It is sufficient that four people show up.
 b. It is sufficient that five people show up.

Here is the semantics of (36a):

- (37) $\exists w': w \sim w'[\neg \exists q[q \in (F) \& q(w') \& \text{that four people show up} \nrightarrow q]]$
 (F) {that four people show up, that five people show up, . . . , that three people show up, . . . }

This says that it is possible that no alternative to 'that four people show up' is true that is not implied by that proposition. Alternatives are propositions 'that n people show up'. If no such alternative that is not implied

EXHAUSTIVITY IN QUESTIONS

by 'that four people show up' is true, then obviously no such alternative that is not implied by 'that five people show up' is true, either. Putting it differently, we predict the entailment pattern for (36) to be like the one in (38):

- (38) It is not necessary that more than four people show up.
 \Rightarrow It is not necessary that more than five people show up.

Sufficient acts as a scale reverser. Its argument is predicted to be a downward monotonic context, in the same sense as in the presuppositional context created by *only*:

- (39) It was sufficient that John sang.
 \Rightarrow It was sufficient that John sang and danced.

- (40) Only John sang.
 \Rightarrow Only John sang and danced.

We have disregarded any presuppositional or implicational semantic contributions of *sufficient*. We do not deny that they are there – they might actually cloud inferential judgements a bit, just as in the case of *only*. We are not certain what the presuppositional contribution is. In the past tense, *be sufficient* + CP seems to presuppose the truth of its complement. We will not go into this any more deeply here.

Getting back to the interrogative that we want to analyze, here is its Hamblin/Karttunen denotation:

- (41) How many eggs are sufficient?
 $\lambda p \exists n [p = \lambda w' [\text{sufficient}(w') \\ (\lambda w'' [\exists X [*egg(w'')(X) \ \& \ \text{card}(X) = n \ \& \ *you_have(w'')(X)])]]]$

Suppose you are OK with four eggs, you don't need more, but you can't have fewer. Then, the set of true propositions in (41) is given in (42):

- (42) {It is sufficient to have four eggs, It is sufficient to have five eggs, . . . }

Suppose that the focus alternatives that *sufficient* is sensitive to are the ones that would be induced by focus on the cardinal number. This is indicated informally in (42'):

- (42') $\{ \lambda w' [\text{sufficient}(w') \\ (\lambda w'' [\exists X [*egg(w'')(X) \ \& \ \text{card}(X) = [4]^F \ \& \ *you_have(w'')(X)])]], \\ \lambda w' [\text{sufficient}(w') \\ (\lambda w'' [\exists X [*egg(w'')(X) \ \& \ \text{card}(X) = [5]^F \ \& \ *you_have(w'')(X)])]], \\ \dots \}$

The inferential behavior of the elements of (42) will then be such that the intersection of these propositions will be the proposition that it is sufficient that you have four eggs, due to the entailment pattern licensed by *sufficient* that we discussed above. Hence, answer1 will be the minimum answer.

4.4. Non-Scalar Predicates and Distributivity

Finally, let's look at a non-scalar question predicate. (43) will receive the interpretation represented in (44).⁴

(43) How many people can play this game?

(44) $\lambda p \exists n [p = \lambda w' [\text{poss}(w') (\lambda w'' [\exists X [\text{people}(w'')(X) \ \& \ \text{card}(X) = n \ \& \ \text{play}(w'')(this_game)(X)]])]]]$

Now suppose that the game can be played with either 4 or 6 people, but no group of any other size can play it. Then the set of true answers to (43), that is, the set of true propositions in (44), will be the set of propositions in (45):

(45) $\{ \lambda w' [\text{poss}(w') (\lambda w'' [\exists X [\text{people}(w'')(X) \ \& \ \text{card}(X) = 4 \ \& \ \text{play}(w'')(this_game)(X)]])], \lambda w' [\text{poss}(w') (\lambda w'' [\exists X [\text{people}(w'')(X) \ \& \ \text{card}(X) = 6 \ \& \ \text{play}(w'')(this_game)(X)]])] \}$

No other propositions will be true; in particular, not the proposition that it is possible for a group of five people to play this game. This is because if there is a group of six people playing this game, there is no proper subgroup of that group that is also playing this game. The predicate *to play this game*, in other words, is not distributive. This causes the difference with the downward scalar case, where we had a distributive predicate, which licenses inferences to subgroups. Intuitions might be even clearer for a predicate like *form a team*, as in (46):

(46) How many people form a soccer team?

Only the proper number of players forms a soccer team – a clear case of a non-distributive predicate. (46) is not that interesting a case in terms of

⁴ Actually, (44) corresponds to one reading of (43) – what might be called the collective reading. There is presumably also a distributive reading, formalized in (i), which is irrelevant for the discussion.

(i) $\lambda p \exists n [p = \lambda w' [\text{poss}(w') (\lambda w'' [\exists X [\text{people}(w'')(X) \ \& \ \text{card}(X) = n \ \& \ *[\lambda x. \text{play}(w'')(this_game)(x)](X)]])]]]$

EXHAUSTIVITY IN QUESTIONS

inferences, since there is only one true answer. Hence, a maximality account doesn't face any problems. But consider (47):

(47) How many people can form a soccer team?

and suppose that you are discussing the different forms in which soccer can be played: indoor, small field,⁵ and regular outdoor. The true answers to (47) are (we think) 6 (indoor soccer), 8 (small field), and 11 (regular). Nothing else is a true answer. And this is predicted by our analysis, since the predicate *form a soccer team* will only apply to groups of 6, 8, and 11 players. The complete answer here is obviously the conjunction of the true propositions, not the maximum answer. Hence, lack of distributivity is one factor that may destroy downward scalarity.

4.5. Conclusion

We have given a detailed analysis of examples of each of the three relevant types of question predicates, and have shown how answer1 is able to capture what counts as a complete answer straightforwardly, without reference to numerical maxima or minima. We believe that the existence of these data shows that what counts as a satisfactory answer to a question must be determined by informativity and inferential properties of the question predicate. We have assumed a particular semantics of *how many*-phrases, and a Hamblin/Karttunen semantics for questions, and shown how this – in interaction with the semantic contribution of other linguistic elements in the question – leads to a certain inferential behavior, and to a prediction about maximally informative answers.

We have only done this here for a few particular examples. Of course, the inferential behavior of predicates is determined by a wide range of factors and this is not the point at which to enter into an extensive discussion of these. A few words might be in order, however, on question predicates that are not downward scalar. For upward scalar predicates, we have discussed a context that makes for a downward monotonic environment: *sufficient*. Clearly, downward monotonic environments in general lead to the upward scalar behavior we observed with *sufficient* (read the indefinite with narrow scope relative to *nobody*):

(48) Nobody solved five problems.
⇒ Nobody solved six problems.

⁵ *Small field* is our translation of German *Kleinfeld*. This is a form of soccer that the first author has had the pleasure to play. She thinks that it is played with 8 people, but isn't quite sure.

We have not presented examples with negative elements because, as is well known, negative elements in degree questions are unacceptable on the relevant narrow scope reading of the indefinite part of the *how many*-phrase (Frampton 1990, Cresti 1995, Rullmann 1995, Beck 1996a,b):

- (49) How many questions did nobody answer?
 For which n: there are n questions that nobody answered.
 #For which n: there is nobody who answered n questions.

Hence, unfortunately, we are more or less robbed of this obvious case of scale reversal. If such questions are acceptable at all, however, we predict that they do receive a minimum interpretation:

- (50) a. Wiewiele Aufgaben haben wenige lösen können?
 how-many problems have few solve can
 'How many problems could few people solve?'
 b.??For which n: few people were able to solve n problems.

And this context licenses upward scalar inferences:

- (51) Few people have been able to solve (as many as) four problems.
 ⇒ Few people have been able to solve (as many as) five problems.

To the extent that (50) is acceptable, it would be another case of scale reversal.

Interestingly, *sufficient* seems exceptional among the downward entailing expressions in that it allows the narrow scope reading of the indefinite part of a *how many*-phrase. At the same time, again in contrast to the other downward monotonic expressions, it is unable to license NPIs. We will not speculate why this is so, but the two facts should probably be related.

A remark on the other data in section 3.1, ex. (16a–d): We have explained the *sufficient* case, and it is fairly clear that all the data we provide there have a similar interpretation. We predict this to the extent that assuming the presence of a scale reversing operator like *sufficient* is plausible in each case.

Now for the non-scalar predicates in section 3.2: we have only discussed effects related to plurality as factors that destroy inferences between groups and subgroups, and hence, cardinalities of groups. There are other contexts in which intuitively, such inferences are not merited, for instance (21a,b), repeated here as (52a,b):

- (52) a. How many courses are you allowed to take?
 b. How high can a helicopter fly?

EXHAUSTIVITY IN QUESTIONS

In (52a), we intuitively get what amounts to an ‘exactly’ interpretation of the cardinal NP: it is allowed to take exactly four courses, or exactly 6, exactly 7, or exactly 8. No other numbers are permissible numbers of courses to take. We cannot derive this effect: any world in which you take at least 6 courses is a world in which you take at least 5 courses. Hence we predict a downward scalar behavior. We don’t know what is responsible for making these examples intuitively non-scalar. (52b) is similar: if we simply assume an ‘at least’ interpretation for degrees, then to fly 200 feet high implies flying 100 feet high, and we do not make the prediction that a complete answer has to specify the minimal altitude as well. Hence we are unable to derive such intuitive non-scalar behavior. We do not understand what goes on in these modal cases and will leave them for future research.

For the three types of degree questions we have looked at, the notion of answer1 gives good results. We will now relate this notion to Rullmann’s original proposal as well as to Groenendijk and Stokhof’s semantics for questions. In the course of doing so we will get back to the issue of strong exhaustivity.

5. STRONG EXHAUSTIVITY AND ANSWER2

5.1. *We Need a Second Notion of Answerhood to Capture Strong Exhaustivity*

Let’s assume that in the actual world John read exactly five books. Now compare (52b), the denotation of answer1 for example (52a), to Rullmann’s (1995) semantics (52c).

- (52) a. How many books did John read?
b. $\lambda w[\text{John read (at least) five books in } w]$
c. $\lambda w[\max(\lambda n[\text{John read } n \text{ books in } w]) = 5]$

The two propositions are not identical. While (52c) contains the information that five is the maximal number of books John read, (52b) expresses just the proposition that John read (at least) five books. Rullmann’s semantics and answer1 also differ in (53), assuming as we did before that in the actual world Mary, Sue, and Jane (and only they) were at the party:

- (53) a. Who was at the party?
b. $\lambda w[\text{Mary} + \text{Sue} + \text{Jane were at the party in } w]$
c. $\lambda w[\max(\lambda x[x \text{ was at the party in } w]) = \text{Mary} + \text{Sue} + \text{Jane}]$

(53c) expresses the proposition that the maximal group that was at the party consists of Mary, Sue, and Jane. (53b) just says that Mary, Sue, and

Jane were at the party, without any information as to whether there were other people there or not. In other words, (53b) gives the complete true answer, while (53c) gives the complete true answer plus the information that this is the complete true answer to the question. In this sense, it includes information about negative instances. Answer1 only incorporates weak exhaustivity, while Rullmann's denotation also captures strong exhaustivity.

Discussion of this issue in the last decade has made clear at least that we need to have strong exhaustivity in some cases, for example in questions embedded under the verb *know* (see below for more extensive discussion). Heim's (1994) paper contains a proposal for how to capture strong exhaustivity on the basis of the original Hamblin/Karttunen denotation. Heim does this by defining a second notion of answerhood, answer2 (again, Q is a question intension):

$$(54) \quad \text{answer2}(w)(Q) = \lambda w'[\text{answer1}(w')(Q) = \text{answer1}(w)(Q)]$$

The answer2 to a question Q is the proposition that the answer1 to Q is what it is in the actual world w . Since this notion of answerhood includes the information that answer1 is the complete answer, it provides information about negative instances, thereby capturing strong exhaustivity. For a non-scalar question like (53a) it strengthens the answer1 given in (53b) by adding that nobody but Mary, Sue, and Jane were at the party. In the case of a scalar question like (52a) it adds the information that (52b) is the maximal true answer, in the sense that John didn't read more than five books.

Answer2 mimics in a very direct way Groenendijk and Stokhof's theory of questions, which was designed to incorporate strong exhaustivity. Heim shows that answer2 is actually equivalent to the denotation Groenendijk and Stokhof assign to interrogatives in most cases, but not in all. She also suggests a way of getting around the problematic cases where the equivalence fails, but we will not go into her solution here. We will show below that full equivalence between answer2 and Groenendijk and Stokhof's question denotation can be proven if we slightly change our assumptions about the original question denotation we start out with. This change involves the distinction between the *de dicto* and *de re* readings of *which*-questions discussed by Groenendijk and Stokhof. But before getting to that we will first briefly introduce Groenendijk and Stokhof's theory of questions.

5.2. Groenendijk and Stokhof (1982, 1984)

Groenendijk and Stokhof (1982, 1984) argue for their semantics for interrogative sentences on the basis of constructions with embedded questions like (55):

EXHAUSTIVITY IN QUESTIONS

(55) John knows who was at the party.

Suppose again that in the actual world w , Mary, Sue, and Jane were at the party and no one else was. Then for (55) to be true, John has to know the proposition that Mary, Sue, and Jane were at the party and no one else was. Groenendijk and Stokhof argue that in w the embedded interrogative denotes exactly that proposition. More generally, in a world w the interrogative (56) denotes the proposition that the set of people at the party is exactly what it is in w . In other words, (56) denotes the set of possible worlds w' such that the set of people at the party in w' is the same as the set of people at the party in w . This denotation is captured more formally by the expression in (57):

(56) Who was at the party?

(57) $\lambda w'[\lambda x[x \text{ was at the party in } w'] = \lambda x[x \text{ was at the party in } w]]$

For John to know (57) means that he has to know that the set of all people who were at the party is what it actually is. This captures strong exhaustivity, since if John mistakenly believed of someone who was not at the party that he was, the set of people at the party would be greater in his belief worlds than in reality, and he would not believe (57).

We will now proceed to show that the notion of answer₂ defined in section 5.1 is equivalent to the denotation Groenendijk and Stokhof assign to interrogatives. We will first do that for the special case of interrogatives with *who* (rather than *which*), for the moment assuming that all entities in our model are people. Later on we will see that the proof generalizes in a trivial way to *which*-questions and to *who*-questions without restrictions on the model, provided we interpret them as de dicto questions in Groenendijk and Stokhof's sense.

What we want to prove is the equivalence of (58) and (59), for arbitrary predicates P of type $\langle s, \langle e, t \rangle \rangle$.

(58) $\lambda w^*[\lambda x[P(w^*)(x)] = \lambda x[P(w)(x)]]$

(59) $\lambda w^*[\cap[\lambda p[\exists x[p = \lambda w'P(w')(x)] \ \& \ p(w^*)]]$
 $= \cap[\lambda p[\exists x[p = \lambda w'P(w')(x)] \ \& \ p(w)]]]$

To make the proof more readable we will follow Heim in using the more convenient set notation in (58') and (59'):

(58') $\lambda w^*[\{x: P(w^*)(x)\} = \{x: P(w)(x)\}]$

(59') $\lambda w^*[\cap\{\lambda w'P(w')(x): P(w^*)(x)\} = \cap\{\lambda w'P(w')(x): P(w)(x)\}]$

$\{x: P(w)(x)\}$ is the set of objects x such that x has property P in world w .

$\{\lambda w'P(w')(x): P(w)(x)\}$ is the set of propositions $\lambda w'P(w')(x)$ such that x has the property P in w .

Proof:

We first show that (58') is a subset of (59'). Take an arbitrary world w^* such that $\{x: P(w^*)(x)\} = \{x: P(w)(x)\}$. We will abbreviate $\{x: P(w^*)(x)\}$ as S^* and $\{x: P(w)(x)\}$ as S . So $S = S^*$. Then:

$$\begin{aligned} & \cap\{\lambda w'P(w')(x): P(w^*)(x)\} = \\ & \cap\{\lambda w'P(w')(x): x \in S^*\} = \\ & \cap\{\lambda w'P(w')(x): x \in S\} = \\ & \cap\{\lambda w'P(w')(x): P(w)(x)\} \end{aligned}$$

Hence if $w^* \in (58')$, then $w^* \in (59')$.

We now do the other half of the proof, namely that (59') is a subset of (58'). This part is a little more complicated; we give a proof by reductio. Suppose that there is a world w^* which is a member of (59'), but not of (58'). Then there must be an individual u such that (i) holds:

$$(i) \quad P(w^*)(u) \ \& \ \neg P(w)(u)$$

(Or the other way around, but that case is completely parallel.) We can then derive the following:

$$\begin{aligned} (ii) \quad & \lambda w'P(w')(u) \in \{\lambda w'P(w')(x): P(w^*)(x)\} && \text{by (i)} \\ (iii) \quad & \cap\{\lambda w'P(w')(x): P(w^*)(x)\} \subseteq \lambda w'P(w')(u) && \text{by (ii)} \\ (iv) \quad & \cap\{\lambda w'P(w')(x): P(w)(x)\} \subseteq \lambda w'P(w')(u) && \text{by (iii) and (59')} \end{aligned}$$

Because of (i) we also know that:

$$(v) \quad w \notin \lambda w'P(w')(u)$$

Finally, the following holds:⁶

$$(vi) \quad w \in \cap\{\lambda w'P(w')(x): P(w)(x)\}$$

We now have a contradiction of (iv), (v), and (vi): $\lambda w'P(w')(u)$ cannot be a superset of $\cap\{\lambda w'P(w')(x): P(w)(x)\}$ if there is a world w that is not in $\lambda w'P(w')(u)$, but is in $\cap\{\lambda w'P(w')(x): P(w)(x)\}$.

End of proof.

⁶ To see why this is the case, take an arbitrary entity v such that $P(w)(v)$. Then $w \in \lambda w'P(w')(v)$. Since this is the case for an arbitrary v , it will be true that $w \in \cap\{\lambda w'P(w')(x): P(w)(x)\}$.

EXHAUSTIVITY IN QUESTIONS

The proof contains one important assumption: that for any interrogative the question predicate *P* is the same in the Groenendijk and Stokhof translation as in the Hamblin/Karttunen translation. For questions with *who* this is obviously the case; however, there is a complication with *which*-questions.

6. DE DICTO VS. DE RE READINGS OF WHICH-PHRASES

6.1. *De Dicto Reading According to Groenendijk and Stokhof and Heim*

Groenendijk and Stokhof point out that there is an ambiguity in embedded interrogatives such as (60).

(60) John knows which students called.

Suppose that in a certain situation the only people who called are Jane, Mary, and Sue, and that John knows this. Suppose furthermore that Jane and Mary are students, but Sue isn't. However, John does not know who the students are, for instance because he mistakenly believes that Sue is also a student. Would we in this scenario be prepared to say that (60) is true? Groenendijk and Stokhof argue that this can be answered both with yes and with no, depending on which reading we assign to (60).

The proposition might be understood so as to not make any claim about whether John knows who is a student and who isn't. If John knows of each individual whether he or she called, then he also knows of each student whether he or she called, and in this sense he knows which students called, even though he may not know whether those who called are students or not. Groenendijk and Stokhof refer to this as the *de re* reading of the *which*-question. However, there is another reading of the question in which for John to know which students called he has to know whether the callers are students. Hence just knowing who called would not imply knowing which students called. This interpretation of the question, which Groenendijk and Stokhof refer to as the *de dicto* reading, can be paraphrased as: John knows of all individuals whether they are students who called.

Groenendijk and Stokhof account for the *de dicto* reading of *which*-phrases in a direct manner. The semantics they would assign to the interrogative *which students called* is given in (61).

$$(61) \quad \lambda w'[\lambda x[\text{student}(w')(x) \ \& \ \text{called}(w')(x)]] \\ = \lambda x[\text{student}(w)(x) \ \& \ \text{called}(w)(x)]]$$

Obviously, knowing this proposition implies knowing what the exact extension of the set of students who called is, and this includes information about

the student status of the callers. The de dicto reading of the whole sentence (60) is represented as:

$$(62) \quad \text{know}(w)(\text{john}, \lambda w'[\lambda x[\text{student}(w')(x) \ \& \ \text{called}(w')(x)]] \\ = \lambda x[\text{student}(w)(x) \ \& \ \text{called}(w)(x)])]$$

Groenendijk and Stokhof take this de dicto reading to be the basic one and derive the de re reading by means of quantifying-in. We'll postpone discussion of that until section 6.4, and for now focus on how to get the de dicto reading in a Hamblin/Karttunen-style semantics.

As Groenendijk and Stokhof point out, Karttunen's (1977) analysis only accounts for the de re reading. Karttunen's translation for the interrogative *which students called* is given in (63).

$$(63) \quad \lambda p \exists x[p(w) \ \& \ \text{student}(w)(x) \ \& \ p = \lambda w'[\text{called}(w')(x)]]$$

This represents the set containing all true propositions of the form 'x called', for x a student. To know which students called, according to Karttunen, means to know all the propositions in this set (or, in our terms, to know the answer1 of (63)). In other words, John knows which students called iff for every student x who called, John knows that x called. This represents the de re reading, since John's knowledge does not include information about whether the callers are students or not. If John knows of both students and non-students who called, then he knows which students called (on this reading), even if he doesn't know who the students are.

Taking Karttunen's de re interpretation as her starting point, Heim (1994) proposes to capture the de dicto reading by means of the notion of answer2. When we apply answer2 to the intension of (63) we get the following:

$$(64) \quad \lambda w^*[\cap[\lambda p \exists x[p(w^*) \ \& \ \text{student}(w^*)(x) \ \& \ p = \lambda w'[\text{called}(w')(x)]]] \\ = \cap[\lambda p \exists x[p(w) \ \& \ \text{student}(w)(x) \ \& \ p = \lambda w'[\text{called}(w')(x)]]]]]$$

Heim shows that (64) is equivalent to Groenendijk and Stokhof's de dicto reading (61), but that the equivalence relies on the assumption that for any two distinct individuals there must be a possible world in which one calls, but the other doesn't. As discussed by Heim, this assumption plausibly holds for 'call', but not for symmetrical predicates like 'live with one's actual spouse'. If x and y are spouses in the actual world, then the worlds in which x lives with x's actual spouse are the same as the worlds in which y lives with y's actual spouse. Heim suggests a solution to this problem in terms of structured meanings, which we will not go into here.

By applying answer2 Heim tries to at the same time account for strong exhaustivity and the de dicto reading. But since the equivalence between

EXHAUSTIVITY IN QUESTIONS

(64) and (61) only holds given certain additional assumptions, this attempt to kill two birds with one stone does not fully succeed. We argue that it is better to keep the two issues separate, and to use *answer2* only to account for exhaustivity. The de dicto reading then has to be derived in a more direct manner. In the next subsection, we develop an empirical argument for this position, based on the semantics of the verb *agree*.

6.2. A Non-Exhaustive De Dicto Reading: The Verb Agree

Our motivation for keeping exhaustivity separate from the issue of de dicto readings is not only that we want to achieve equivalence of *answer2* with the Groenendijk and Stokhof denotation without special assumptions about the predicates involved. There also are empirical reasons to believe that the existence of a de dicto reading of *which*-phrases is independent of exhaustivity. Certain questions permit a reading which is de dicto but at the same time non-exhaustive. A case in point are questions embedded under the verb *agree (on)*, which is discussed at length by Lahiri (1991). We will not go too deeply into his discussion, which centers around quantificational variability effects; those tend to complicate matters considerably. Here, we will ignore these complications, although it is of course clear that those cases have to be looked at. We will focus on a property of *agree* which is very rare for question-embedding predicates: it does not require the answer-propositions to the question it embeds to be true. For instance, in (65),

(65) John and Bill agree on who was elected.

the propositions that John and Bill agree on do not have to be true; that is, they have to be possible answers to the question rather than true answers. Notice that this alone suffices to make the standard notions of exhaustivity inapplicable. For John and Bill to agree on Q, they need not bear any specific propositional attitude to the set of true answers to Q. They could be completely mistaken, or fail to have any beliefs about actual true answers to Q, and still agree on Q. Hence *agree* does not make use of either *answer1* or *answer2*. (*Agree* is also the main reason why we modified the Karttunen semantics by eliminating the requirement that the propositions in the question denotation be true, adopting the Hamblin/Karttunen denotation as the basic question meaning.)

We now suggest the following semantics for *agree on* (which would probably have to be modified to account for quantificational variability data, but should suffice for our purposes):

- (66) $\text{agree}(w)(X)(Q)$ iff for all x in X , all p in Q :
- (i) if x believes p in w , then for all y in X : y believes p in w ; and
 - (ii) if x believes not- p in w , then for all y in X : y believes not- p in w .

where Q is the Hamblin/Karttunen denotation.

Agree itself is thus a counterexample to exhaustivity: Since it does not make use of true answers but of possible answers, there is no meaningful way in which it can be said to be exhaustive. At the same time, we think *agree* should not be characterized as intensional in Groenendijk and Stokhof's sense (which we think would be their way of dealing with *agree*): the lexical semantics of *agree* makes use of answers to the embedded question, and thus of identifiable propositions.

Now what about *agree on* and *which*-questions? A de dicto reading of a *which*-phrase such as *which physicists* would imply that the individuals denoted by the subject of *agree* have to also agree on the physicist status. Hence for (67) to be true, people would have to be physicists in John and Bill's belief worlds rather than the actual world. On such a reading, (67) would be true and appropriate in the scenario described in (68).

- (67) John and Bill agree on which physicists are Nobel prize winners (this year).
- (68) Scenario: Agatha and Dorothy are physicists and Nobel prize winners. Waldemar is a Nobel prize winner, but not a physicist. Pumuckl is neither. John and Bill believe that Agatha, Waldemar, and Pumuckl are physicists and Nobel prize winners. They have no beliefs about the others.

An example that illustrates the existence of such a reading quite clearly is one in which the subjects of *agree* are likely to be mistaken about the restriction. Consider (69):

- (69) Jonas and Ida agree on which European countries have a king.

Jonas and Ida are small children. Intuitively, (69) can be true if Jonas and Ida believe of the same entities that they are European countries and kingdoms, even though the entities are not European countries at all. This is very clearly a de dicto reading. As we said above, *agree on* does not seem to be amenable to an analysis in terms of exhaustivity at all. Hence we have shown that there are non-exhaustive de dicto readings. We conclude that de dicto readings should be dissociated from the issue of exhaustivity,

EXHAUSTIVITY IN QUESTIONS

and that therefore we should not rely on answer2 (which will yield a strongly exhaustive interpretation) to produce the de dicto reading, as Heim does.

6.3. Accounting for the De Dicto Interpretation in the Basic Question Denotation

We have argued that a de dicto interpretation of *which*-phrases should be incorporated into the basic Hamblin/Karttunen semantics. Note that the proof we gave above for the equivalence of (58) and (59) does not apply to (61) and (64). The reason is that the question predicates (i.e., what corresponds to P in the proof) differ in the two formulas. In (61) the question predicate is the property of being a student who called, but in (64) it simply is the property of having called. This is the key to finding a solution to the problem of accounting for the de dicto reading. We propose to give (70a) rather than (70b) as the basic translation for the interrogative *which students called*.

- (70) a. $\lambda p \exists x [p = \lambda w' [\text{student}(w')(x) \ \& \ \text{called}(w')(x)]]$
 b. $\lambda p \exists x [\text{student}(w)(x) \ \& \ p = \lambda w' [\text{called}(w')(x)]]$

Speaking informally, what we are proposing is to make the property denoted by the N' sister of *which* (i.e., *student* in the present example) part of the content of the propositions contained in the denotation of the question, rather than treating it as an 'external' condition determining which propositions go into that set. In fact, this is not a new proposal. It is in line with Hamblin's (1973) original analysis as opposed to Karttunen's, abstracting away from the difference in theoretical framework. For more extensive discussion, see Rullmann and Beck (1998, to appear).

The answer2 of the intension of (70a) is this:

- (71) $\lambda w * [\cap [\lambda p [\exists x [p = \lambda w' [\text{student}(w')(x) \ \& \ \text{called}(w')(x)]]] \ \& \ p(w*)]]$
 $= \cap [\lambda p [\exists x [p = \lambda w' [\text{student}(w')(x) \ \& \ \text{called}(w')(x)]]] \ \& \ p(w)]]$

Now the question predicates in (71) and in the de dicto Groenendijk and Stokhof translation (61) are identical: both are the property of being a student who called. This means that the proof we gave above for the equivalence of answer2 and the Groenendijk and Stokhof denotation will now go through; that is, (71) is equivalent to (61) (just substitute the property $\lambda w \lambda x [\text{student}(w)(x) \ \& \ \text{called}(w)(x)]$ for P in the proof).

Recall that for our proof of the equivalence of answer2 with Groenendijk and Stokhof's semantics in the case of *who*-questions, we made the assumption that the model only contain entities that are persons. We can now safely

drop this assumption if we treat *who* as if it was *which person*. The de dicto interpretation of the question *Who called?* will then be as follows:

$$(72) \quad \lambda p \exists x [p = \delta w' [\text{person}(w')(x) \ \& \ \text{called}(w')(x)]]$$

Before we proceed we would like to add a cautionary note. Our discussion of the semantic contribution to question meaning of the common noun in the *which*-phrase has been rather superficial. There are some important problems with treating the common noun in the way we've done in this paper. We discuss these in Rullmann and Beck (1998, to appear), where we argue that the common noun has the semantic status of a presupposition. We also provide additional arguments for interpreting the common noun inside the propositions that make up the question meaning. The results we discussed above concerning the equivalence of answer2 to the Groenendijk and Stokhof interpretation will carry over to the presuppositional treatment in an appropriate way (Rullmann and Beck 1998).

6.4. Deriving the De Re Readings

Groenendijk and Stokhof propose to account for the de re reading of (60) (= *John knows which students called*) by quantifying-in the common noun of the *which*-phrase at the level of the matrix clause. This yields the following translation:

$$(73) \quad \text{know}(w)(\text{john}, \lambda w' [\lambda x [\text{student}(w)(x) \ \& \ \text{called}(w')(x)] \\ = \lambda x [\text{student}(w)(x) \ \& \ \text{called}(w)(x)]])$$

We will not go into the details of the quantifying-in mechanism here, because we do not want to discuss the compositional derivation of these readings in any detail. The difference between the de dicto reading (62) and the de re reading (73) resides in the world variable of the predicate 'student'. On the de dicto reading (62), John knows the proposition which is true in all those worlds w' in which the set of students who called is the same as the set of students who called in the actual world w . In the de re reading (73), on the other hand, John knows the proposition which is true in all and only the worlds w' such that the set of individuals who were students *in the actual world* w and who called in w' is the same as the set of individuals who were students in w and called in w .

We can capture the de re/de dicto ambiguity by allowing freedom of choice for the possible-world variable on the N' of the *which*-phrase (i.e., *students* in (60)), that is, by varying what we substitute for the blank in (74). (For arguments for a free choice of the world variable in nominal

EXHAUSTIVITY IN QUESTIONS

categories see Enç (1986) and Heim (1991). The details of what exactly is permissible remain to be worked out.)

$$(74) \quad \lambda p \exists x [p = \lambda w' [\text{student}(_)(x) \ \& \ \text{called}(w')(x)]]$$

Choosing the variable w' , which is locally bound by the λ -operator, yields the de dicto reading of the complement represented in (70a). As shown in the previous subsection, when we apply *answer2* to the resulting intension of (70a) we get a reading that is equivalent to Groenendijk and Stokhof's de dicto interpretation (61).

The de re reading is derived by choosing a different variable than w' . It is actually not quite accurate to speak of *the* de re reading, because in many cases there is a choice among several non-locally bound variables as the possible-world argument for the noun *students*. Let's assume for the moment that the verb *know* selects as its argument the strongly exhaustive *answer2*, rather than the basic Hamblin/Karttunen denotation (see below for discussion). To get Groenendijk and Stokhof's de re reading, we have to identify the world variable on the noun *students* with the world with respect to which the sentence as a whole is evaluated (i.e., the actual world). This gives us the following translation:

$$(75) \quad \text{know}(w)(\text{john}, \text{answer2}(w)(\lambda w'' \lambda p \exists x [p(w'') \ \& \ p = \lambda w' [\text{student}(w)(x) \ \& \ \text{called}(w')(x)]]))$$

By the definition of *answer2*, (75) equals (76).

$$(76) \quad \text{know}(w)(\text{john}, \lambda w^* [\cap [\lambda p [\exists x [p = \lambda w' [\text{student}(w)(x) \ \& \ \text{called}(w')(x)]] \ \& \ p(w^*)]] \\ = \cap [\lambda p [\exists x [p = \lambda w' [\text{student}(w)(x) \ \& \ \text{called}(w')(x)]] \ \& \ p(w)]]])$$

(76) is in turn equivalent to Groenendijk and Stokhof's de re reading (73). This equivalence is just a special instance of that between (58) and (59), which was proven in section 6.2 (fill in ' $\lambda w' \lambda x [\text{student}(w)(x) \ \& \ \text{called}(w')(x)]$ ' for the predicate P).

By identifying the world variable on the head noun of the *which*-phrase with different variables, we can thus derive different readings for the question. If this world variable is bound locally, we get the de dicto reading. If it is bound non-locally, we get various sorts of de re readings. By adding further levels of embedding, further possibilities arise. In *Sheila believes that John knows which students called*, for instance, it makes a difference whether we identify the index on *students* with the actual world or with Sheila's belief worlds. The readings we obtain in this way correspond exactly to the readings that are generated in Groenendijk and Stokhof's

theory by applying quantifying-in to the head noun of the *which*-phrase at different levels.

In conclusion, we have shown that by employing answer2 plus a freedom of choice for the world variable on the head noun of *which*-phrases, we can generate exactly the same family of strongly exhaustive readings for interrogatives that Groenendijk and Stokhof do. This raises the question of why we went through all this trouble of defining notions of answerhood on the basis of a Hamblin/Karttunen semantics, if we can obtain the same family of strongly exhaustive interpretations in a Groenendijk and Stokhof semantics straightforwardly without such a fuss. The answer to this question (which will be elaborated in the next two sections) is that there are cases in which questions are interpreted non-exhaustively and in which we therefore do not want to apply answer2 to the basic (Hamblin/Karttunen) question denotation. To deal with such cases, a richer system of potential question interpretations is called for, including weakly exhaustive and non-exhaustive ones.

7. ARGUMENTS FOR A FLEXIBLE APPROACH TO EXHAUSTIVITY

Summarizing our discussion so far, we have now arrived at a theory of questions which makes available at least three distinct semantic objects that are associated with a question. First, there is the Hamblin/Karttunen denotation, the set of all propositions that count as (not necessarily exhaustive) answers to the question. Let's call this set $Q(w)$ (Q being the Hamblin/Karttunen intension). Second, we have $\text{answer1}(w)(Q)$, the proposition that is the intersection of all the true members of $Q(w)$. This constitutes the weakly exhaustive true answer to the question. Third, we have $\text{answer2}(w)(Q)$, which is the strongly exhaustive answer to the question and which is the same as the denotation that Groenendijk and Stokhof assign to questions. (In *which*-questions additional *de re* readings are possible, but in the remainder of this paper we will leave those aside.) An important question that arises then is whether we really need all three of these notions. Couldn't we just use the strongly exhaustive answer2, as Groenendijk and Stokhof have argued so forcefully throughout their work?

We have in fact already seen one strong empirical argument against treating all questions as exhaustive: the semantics of the verb *agree*, whose complement is not even weakly exhaustive because it must include false answers (see section 6.2). In this section we will discuss further evidence against the idea that questions are always exhaustive, arguing that having all three notions of answerhood allows us to adopt a more flexible theory that takes into account cases in which insisting on strong exhaustivity

EXHAUSTIVITY IN QUESTIONS

gives rise to truth conditions that appear to be stronger than is supported by our intuitions. In taking this position we have been inspired by Heim (1994), who also provides many of the arguments discussed in this section.

Our discussion is primarily intended to motivate a flexible approach to exhaustivity, rather than to argue against a specific alternative like Groenendijk and Stokhof's. The points we are going to make are not necessarily problematic for Groenendijk and Stokhof when taken individually – Groenendijk and Stokhof explicitly discuss and account for some of them, in particular the 'mention-some' interpretation. We nonetheless think that the global picture that emerges supports a rich and flexible system which provides a range of interpretations for questions with various degrees of exhaustivity, because of its greater overall simplicity and elegance. In addition, at certain points we will present facts that we do judge problematic for Groenendijk and Stokhof, or in fact any theory that treats strong exhaustivity as a property of the basic question denotation. We will mention that explicitly in each case.

Maybe at this point a remark on the status of these notions of answerhood is in order. It might be supposed that they are convenient technical notions designed to give the right results formally, but without much conceptual foundation. We think that the opposite is the case. We consider Heim's formalization of *answer1* and *answer2* an important step in developing an understanding of the role and interpretation of interrogatives in natural language semantics. The idea is that interrogatives do not always enter semantic composition with their basic question denotation. To give an example, if 'x knows Q' is true (where Q is an interrogative sentence), then the subject does not stand in the 'know'-relation to a question (in contrast, possibly, to relations like 'wonder', which might be a relation a subject bears towards a question intension), but to the answer, in some sense, to the question. Since for 'x knows Q' to be true, x has to know a certain proposition, Groenendijk and Stokhof conclude that Q denotes that proposition. Heim's proposal is quite different: one does not know a question – one knows an answer to a question. Therefore, we need not conclude that Q itself denotes a proposition. Rather, it seems a natural mechanism of the interpretation of interrogatives to step from the question denotation to an answer to that question.

The notion of an answer to a question leaves room for various concepts of what constitutes an answer. This will also crop up in the discussion of the noun *answer*. *Answer1* and *answer2* are two possible formal notions that can be employed, both corresponding to intuitive concepts of complete answers. Later on we will see that there is at least one more concept of

answerhood, namely the 'simple' answers constituted by the individual propositions in the Hamblin/Karttunen denotation.

7.1. *Mention-Some Readings*

One argument showing that sometimes questions are not even weakly exhaustive can be based on what Groenendijk and Stokhof call the 'mention-some' interpretation of questions (see especially Groenendijk and Stokhof 1984, chapter 6). Some examples which favor the mention-some interpretation are the following:

- (77) a. John knows where you can buy the New York Times.
 b. Mary told me how to get to the train station.

Sentence (77a) for instance has a reading on which it is true even if John isn't able to provide a complete list of places where one can buy the New York Times, but only one particular location, say, the newsstand at the train station. Groenendijk and Stokhof account for the existence of the mention-some interpretation in terms of disjunctions of questions – an analysis which we won't discuss here. What is relevant for our present purposes is that the mention-some interpretation can be straightforwardly captured if we can avail ourselves of the Hamblin/Karttunen denotation of the embedded question. On the mention-some interpretation, (77a) will be true iff John knows at least one of the propositions in the Hamblin/Karttunen denotation of the embedded question which is true in the actual world. This reading is represented in (78).

- (78) $\exists p[\text{know}(w)(\text{john}, p) \ \& \ \text{where-can-you-buy-the-NYT}'(w)(p) \ \& \ p(w)]$

In section 8 we will make two separate proposals about how this reading could be derived. One relies on a lexical rule applying to the matrix verb, the other on a type-shift applying to the complement question.

7.2. *Surprise, Predict, and Other Verbs*

As Heim points out, although it is possible to define answer2 in terms of answer1 and answer1 in terms of the Hamblin/Karttunen denotation, this is crucially a one-way street. When we have only answer2 (or in fact any strongly exhaustive question interpretation), it is not possible to get back answer1 or the Hamblin/Karttunen denotation. In a certain sense, answer2 contains less information than answer1 and the Hamblin/Karttunen denotation. For instance, because a question and its negation impose the same

EXHAUSTIVITY IN QUESTIONS

partition on the set of possible worlds, (79a) and (79b) will have the same answer₂, but their Hamblin/Karttunen denotation and answer₁ will generally differ:⁷

- (79) a. Who was at the party?
 b. Who was not at the party?

One potential problem for strong exhaustivity will therefore be question-embedding verbs which discriminate between an embedded question and its negation.⁸ A case in point are emotive factives such as *surprise* (Berman 1991, Heim 1994). (80b) may very well be true although (80a) is false, for instance in a situation in which everyone who was at the party was expected to be there by the speaker, but some people who were also expected to be there did not show up.⁹

- (80) a. It surprised me who was at the party.
 b. It surprised me who was not at the party.

⁷ To see this, suppose that Mary, Sue, and Jane were at the party and Bill, Graham, and Marc were not, and that they are all the people in the context. Then answer₁(w)((79a)) will be (i), and answer₁(w)((79b)) will be (ii).

- (i) $\lambda w[\text{Mary} + \text{Sue} + \text{Jane were at the party in } w]$
 (ii) $\lambda w[\text{Bill} + \text{Graham} + \text{Marc were not at the party in } w]$

Answer₂(w)((79a)) will be (iii), and answer₂(w)((79b)) will be (iv):

- (iii) $\lambda w[\text{answer}_1(w)((79a)) = \lambda w'[\text{Mary} + \text{Sue} + \text{Jane were at the party in } w']]$
 (iv) $\lambda w[\text{answer}_1(w)((79b)) = \lambda w'[\text{Bill} + \text{Graham} + \text{Marc were not at the party in } w']]$

However, these two sets of possible worlds will be identical, since whenever the answer₁ to (79a) will be the proposition that Mary, Sue, and Jane were at the party, the answer₁ to the negated question will be that the complement of Mary, Sue, and Jane in the universe of discourse were not at the party.

⁸ Groenendijk and Stokhof (1989) are aware that the equivalence of a question with its negation on a theory incorporating strong exhaustivity is problematic. They discuss the issue in the context of the relation between matrix questions and term answers. The term answer *Mary, Sue, and Jane* expresses a different proposition when it is an answer to (79a) than when it is an answer to (79b). They point out that the problem does not arise in the categorial analysis of questions which treats questions as unsaturated objects (n-place predicates) rather than (sets of) propositions, and suggest that the propositional and the categorial approach can be unified within a type shifting framework, a proposal which appears to be compatible with our own type shifting approach in section 8.2. Groenendijk and Stokhof do not seem to view the predicted equivalence of questions with their negations as a problem when it comes to *embedded* questions.

⁹ We believe that this issue is independent of the speaker's knowledge of what is included in the domain, discussed by Groenendijk and Stokhof as a potential confounding factor; that is, (80a) and (80b) are not necessarily equivalent even if the matrix subject (in this case the speaker) is fully informed as to what the entities included in the domain of discourse are.

This can be captured easily if we have the notion of answer1 at our disposal, but not if the only thing we have is answer2.

Another question-embedding predicate that supports the conclusion that having only a strongly exhaustive question denotation is not enough is *predict*.¹⁰ Take (81), which we modelled on the attested example (82):

(81) I was better at predicting who would show up than I was at predicting who wouldn't show up.

(82) Het blijkt overigens gemakkelijker te voorspellen
it turns-out by-the-way easier to predict
wie niet moe zal worden dan wie wel moe
who not tired will become than who PRT tired
wordt. (De Volkskrant, 10/2/1995)
becomes.

'It turns out it is easier to predict who will not become tired than who will.'

Let's consider for a moment what (81) means. Suppose that you predicted of 10 people that they would show up, and 8 of them actually did, and you also predicted of 10 people that they would not show up, and 5 of them actually didn't. Hence you predicted with more accuracy the true answers to the affirmative question than to the negative question. This is the sort of scenario that we take (81) to describe.

To discuss the example properly, we first need to make up our minds as to what *predict* means in a case like (82):

(82) I predicted 90% accurately who would show up.

Suppose that there are 10 people such that you predicted that those people would show up. Nine of them did. We take (82) to be a true description of that scenario.

Given that we assume a Hamblin/Karttunen denotation for 'who would show up', this means that we have to count the true propositions in that set of propositions which are about atomic individuals. Those are the propositions that do not imply any of the other propositions in that set, since, e.g., 'Cecile + Gereon showed up' implies that Cecile showed up, but the proposition that Cecile showed up does not imply any other proposition in the Hamblin/Karttunen denotation.

¹⁰ Special thanks once more to Irene Heim, to whom we owe the idea for the argument made in this section.

EXHAUSTIVITY IN QUESTIONS

So, *predict* has to have a semantics something like (83):

$$\begin{aligned}
 (83) \quad & \text{predict}(w)(Q)(x)(n\% \text{ accurately}) = 1 \text{ iff} \\
 & \text{card}(\{p: \text{predict}(w)(p)(x) \ \& \ Q(w)(p) \ \& \ p(w) \ \& \\
 & \quad \neg \exists q [Q(w)(q) \ \& \ p \subseteq q]\}): \\
 & \text{card}(\{p: \text{predict}(w)(p)(x) \ \& \ Q(w)(p) \ \& \\
 & \quad \neg \exists q [Q(w)(q) \ \& \ p \subseteq q]\}) \\
 & = n: 100
 \end{aligned}$$

That is, to predict Q n percent accurately, n percent of the atomic propositions in Q that you predict have to be true. We reduce *predict* + question to *predict* + proposition. This semantics implies that we have to be able to count atomic propositions in the Hamblin/Karttunen denotation. Now we can make sense of (81): The Hamblin/Karttunen denotations of ‘who showed up’ and ‘who didn’t show up’ are different, and contain different atomic propositions. That means that you can correctly predict a greater percentage of the former than of the latter. We take (81) to be true if your accuracy of prediction as defined in (83) was higher in the case of the affirmative question than it was in the case of the negated question.

If we only had a strongly exhaustive question denotation at our disposal, the two questions would have the same denotation, as discussed in the *surprise* case above. We could not explain that (81) is non-contradictory. Notice also that this does not seem to be a case of a mention-some reading. Hence once more, we have found support for a richer universe of question denotations. Notice also that if the above semantics of *predict* is correct, we have to include non-true answers in the question set, which is an additional reason for starting with the Hamblin/Karttunen denotation.

Counting propositions in the question denotation should remind the reader strongly of cases of so-called quantificational variability, as discussed by Berman (1991) and Lahiri (1991). It may be argued that a similar counting of atomic propositions goes on in an example like (84):

(84) I was mostly surprised by who showed up.

We believe that this means that for most people who in fact showed up, you were surprised that that person showed up. This is the semantics predicted by Berman. We suggest that this might be captured by quantifying over true atomic propositions in the question denotation, as demonstrated for *predict* above: You were surprised at more of those propositions than there were such propositions that you weren’t surprised at. This is contrary to Berman, who quantifies over individuals to capture this reading. It is also not the same as Lahiri’s analysis, which introduces a lattice structure over the propositions in the question set. We believe that (84) obligatorily

distributes over atomic propositions – it can't express surprise at particular group show-ups – and hence we see no need for a lattice structure; in fact we think that it is undesirable to have one. Be that as it may – we won't discuss quantificational variability in detail here, and only want to bring examples like (84) to the reader's attention since, as Berman (1991) observes, they also show that questions do not uniformly have a strongly exhaustive interpretation. The argument in the case of (84) (whatever will ultimately turn out to be the best analysis) is that we only want to count propositions about people who did show up, not also the ones about people who didn't show up: as Berman observes, (85a) does not imply (85b), which it would given a strongly exhaustive interpretation of the embedded question.

- (85) a. John mostly knows who ran.
 b. John mostly knows who did not run.

Another set of examples illustrating a similar problem are propositional attitude verbs which refer to ways of conveying information. The verbs we have in mind include *tell*, *read*, *write down*, *list*, and *enumerate*. These verbs seem to have two distinct senses (although it's not at all clear whether we are dealing with an actual ambiguity here), a transparent and a non-transparent one (cf. Heim 1994). On the transparent sense of *read*, reading who was at the party implies reading who was not at the party. This is the sense that Groenendijk and Stokhof seem to have in mind when they argue for strong exhaustivity. However, although we acknowledge the existence of the transparent sense, we believe that there also is a sense in which (86a) and (86b) are not equivalent – and in fact we tend to think that this is the sense in which this class of verbs is ordinarily understood:

- (86) a. John read/wrote down who was at the party.
 b. John read/wrote down who was not at the party.

We agree with Heim (1994) that there is an interpretation for these verbs (which is still fairly transparent, but not completely) in which for instance *tell* means something like 'cause one to know answer2 by asserting answer1'. The point is that in this sense the verbs would make use of answer1 as well as answer2. Such cases are not restricted to artificially constructed data. We came across the following real-life example in a recent article in *Language* discussing the occurrence of affixes of a certain form among native American languages:¹¹

¹¹ J. Nichols and D. A. Peterson, 'The Amerind Personal Pronoun', *Language* 72, 336 (1996). The cited sentence is a quotation from Greenberg.

EXHAUSTIVITY IN QUESTIONS

- (87) In fact, it would probably be easier to enumerate where *nV*- and *mV*- are not found than where they are.

To give this sentence a noncontradictory reading it is necessary to assign different denotations to the embedded questions where *nV*- and *mV*- are found and where *nV*- and *mV*- are not found.¹²

We believe that examples of this type involve answer1 but not answer2, and hence these predicates can be true of a question without being true of its negation. The questions are weakly but not strongly exhaustive.

7.3. *The Noun Answer*

Another type of case in which answer1 seems to play a role in embedded *wh*-constructions involves the semantics of the noun *answer* (Heim 1994). Heim notes that (88) may be true in a situation where John just happens to know a proposition which constitutes the weakly exhaustive answer to the embedded question, even if he is not aware that it is the weakly exhaustive answer.

- (88) John knows the answer to the question who was at the party.

Suppose that Mary and Sue were the only party guests. Then (88) is true if John knows the proposition that Mary and Sue were at the party, even if he believes (wrongly) that others attended the party as well. The noun *answer* must therefore mean answer1. But because answer1 cannot be retrieved from answer2, this implies that the embedded question itself cannot be strongly exhaustive.

Heim's argument can actually be extended to show that in certain cases the noun *answer* is not even weakly exhaustive. This is the case when it is combined with an indefinite determiner as in (89):

- (89) John knew only one answer to the question which Dutch Olympic athletes won a medal.

This sentence will be true iff John was able to mention exactly one Dutch Olympic medalist. In this case the noun *answer* must be construed as referring to non-exhaustive true answers, that is, true propositions in the Hamblin/Karttunen denotation. This concept of answerhood will be formalized later. It is an instance of the so-called 'mention-some' reading

¹² Martin Stokhof (p.c.) has suggested to us that to deal with examples with *list* or *enumerate* we may need recourse to the categorial analysis of questions, because what is listed or enumerated are often term answers rather than propositional answers. See also footnote 9.

of questions, which will be reanalyzed as involving a third notion of answerhood. Thus we see that intuitions about the noun *answer* can be regarded as the intuitive foundation of the formal definitions of answerhood. Note also that in (89) the head noun *Dutch Olympic athletes* can be read de dicto. This means that this sentence provides additional evidence for the existence of questions with a non-exhaustive de dicto reading. As noted in section 6.2, such a reading is predicted not to be possible in Heim's (1994) account, which relies on the strongly exhaustive answer² to derive the de dicto reading. It shows that being read de dicto is a property of *which*-phrases in the basic question denotation.

7.4. (Non-)Exhaustivity Markers

Yet another argument in favor of an approach which recognizes the existence of non-exhaustive readings derives from the use of various linguistic expressions to explicitly mark a question as being understood either exhaustively or non-exhaustively. One such marker is the expression *for example* in (90):

- (90) Who for example was at the party last night?

By adding *for example* the speaker makes explicit that she will be satisfied with a non-exhaustive (in the weak sense) answer to the question. *For example* cannot easily occur in embedded questions. However, there are other non-exhaustivity markers in other languages that can. In Dutch we find *zoal* (see (91)), and in German its cognate *so* (see (92)).

- (91) Jan wil weten wie er zoal (niet) op het feest
 Jan wants know who there ZOAL (not) at the party
 waren.
 were

'John wants to know who for example were (not) at the party.'

- (92) Hans will wissen, wer so (?nicht) auf dem Fest
 Hans wants know who SO (not) at the party
 war.
 was

'John wants to know who for example were (not) at the party.'

What these sentences express is that Jan/Hans wants a few representative examples of people who were at the party. While this interpretation is a rather vague one, it is clearly non-exhaustive. Although we will not provide

EXHAUSTIVITY IN QUESTIONS

a formal interpretation for non-exhaustivity-markers, we believe that intuitively their existence shows that questions in natural language in principle have the option of being interpreted non-exhaustively. A question can be marked to enforce this interpretation.

In addition to non-exhaustivity markers we find expressions in natural language which can be used to indicate exhaustivity. German has the word *alles*, which has exactly that function (Reis 1992, Beck 1996a), and in Dutch we find the corresponding *allemaal*:¹³

(93) Hij weet wie er allemaal op het feest waren.
he knows who there all at the party were
'He knows who all were at the party.'

(94) Er weiss, wer alles auf dem Fest war.
he knows who all on the party was
'He knows who all were at the party.'

What these expressions do is force a (weakly) exhaustive interpretation of the question in which they are contained. They are incompatible with a mention-some interpretation:

(95) Hans weiss wo man alles/überall die
Hans knows where one all/everywhere the
New York Times kaufen kann.
New York Times buy can
'Hans knows where all you can buy the New York Times.'

In contrast to (77a), (95) does not have the mention-some interpretation and can only be interpreted exhaustively. It should be pointed out though that *alles* and *allemaal* do not force strong exhaustivity, which explains why they are not incompatible with the class of verbs mentioned earlier like *surprise*, which disprefer a strongly exhaustive interpretation:

(96) Es hat mich überrascht, wer alles auf dem Fest war.
it has me surprised who all at the party was
'It surprised me who all was at the party.'

¹³ To these examples from our own native languages we could add (i) from a dialect of Irish English discussed by McCloskey (1995):

- (i) a. What all did you get for Christmas?
b. What did you get all for Christmas?

We suggest the following semantics for *alles/allemaal*:

$$(97) \quad \text{alles}(w)(Q) = \lambda p[p = \text{answer1}(w)(Q)]$$

Alles operates on a question denotation and gives us the weakly exhaustive interpretation, that is, the singleton set containing *answer1*. Since we propose to deal with mention-some interpretations via the elements in the Hamblin/Karttunen denotation, from (97) there will be no way back to a real mention-some interpretation. The only element in the set denoted by the question is already weakly exhaustive.

We believe that just like non-exhaustivity markers such as *for example*, exhaustivity markers like German *alles* pose a challenge to a theory that uniformly gives every question an exhaustive interpretation. If the basic meaning of questions already were an exhaustive one, exhaustivity markers would be superfluous and the question with the exhaustivity marker should have exactly the same range of interpretations as the corresponding question without it. However, we feel that this is not the case: (95) differs in meaning from (77a) in that the former does not allow a non-exhaustive interpretation whereas the latter does. A rigid approach to exhaustivity will have no way to deal with this difference (for instance Groenendijk and Stokhof's (1984) approach to the mention-some interpretation could not, in our judgment, predict that (95) does not have a mention-some interpretation, since *alles* could make no difference to the original question interpretation).

7.5. Degree Questions with At Least/At Most

The next argument gets us back to the issue of degree questions. Consider the paradigm in (98).

- (98) a. *Wieviele Leute waren da?*
 how-many people were there
 'How many people were there?'
- b. *Wieviele Leute waren mindestens da?*
 how-many people were at-least there
 'How many people were there at least?'
- c. *Wieviele Leute waren höchstens da?*
 how-many people were at-most there
 'How many people were there at most?'

The intuition is clear that (98a–c) each mean something different. This holds also for the embedded case:

EXHAUSTIVITY IN QUESTIONS

- (99) a. Hans weiss, wieviele Leute da waren.
 Hans knows how-many people there were
 'Hans knows how many people were there.'
- b. Hans weiss, wieviele Leute mindestens da waren.
 Hans knows how-many people at-least there were
 'Hans knows how many people were there at least.'
- c. Hans weiss, wieviele Leute höchstens da waren.
 Hans knows how-many people at-most there were
 'Hans knows how many people were there at most.'

(99b) and (99c) are actually a bit odd. We will come to a tentative explanation for that in a moment.

For (99a) to be true, Hans has to know the exact number of people who were there. For (99b) to be true, he has to know a reasonable lower bound of the number of people who were there, for (99c) a reasonable upper bound. So for example, if in fact 86 people were there, and Hans knows that definitely no more than 90 people were there, one could truthfully utter (99c).

For the following formal discussion we will assume that *at least* and *at most* mean exactly what they normally do, namely (100).

- (100) a. at least n (N) (P) iff $\text{card}(\lambda x[N(x) \ \& \ P(x)]) \geq n$
 b. at most n (N) (P) iff $\text{card}(\lambda x[N(x) \ \& \ P(x)]) \leq n$

(101a–c) are the Karttunen denotations of (98a–c).

- (101) a. $\lambda p \exists n[p(w) \ \& \ p = \lambda w'[\exists x[\text{people}(w')(x) \ \& \ \text{card}(x) = n \ \& \ \text{were-there}(w')(x)]]]$
 b. $\lambda p \exists n[p(w) \ \& \ p = \lambda w'[\text{card}(\lambda x[\text{people}(w')(x) \ \& \ \text{were-there}(w')(x)]) \geq n \]]$
 c. $\lambda p \exists n[p(w) \ \& \ p = \lambda w'[\text{card}(\lambda x[\text{people}(w')(x) \ \& \ \text{were-there}(w')(x)]) \leq n \]]$

Now suppose that in w 86 people were there. Then (101a–c) denote the sets given in (102a–c).

- (102) a. {86 people were there, 85 people were there, . . . }
 b. {at least 86 people were there, at least 85 people were there, at least 84 people were there, . . . }
 c. {at most 86 people were there, at most 87 people were there, at most 88 people were there, . . . }

Applying answer1 to these sets will result in the following propositions:

- (103) a. $\lambda w[86 \text{ people were there in } w]$
 b. $\lambda w[\text{at least } 86 \text{ people were there in } w]$
 c. $\lambda w[\text{at most } 86 \text{ people were there in } w]$

This would mean that for Hans to know how many people were there at most, he would have to know that at most the actual number of people were there. This is not the result we intuitively want: It is sufficient for Hans to know that definitely no more than a reasonable upper bound of the actual number of people were there. The same holds for (99b). The ordinary Groenendijk and Stokhof interpretation runs into the same problem.¹⁴

What is going on here? We think that in (99b,c) the mention-some interpretation is the only one that makes sense. An exhaustive interpretation of any kind will always lead to unintuitive results in that the resulting interpretation predicts truth conditions that are too strong. So technically (98b,c) and (99b,c) are just more instances of a mention-some interpretation. We have discussed them separately because (i) the data are quite interesting by themselves, and (ii) they show that non-exhaustivity in the case of degree questions will be non-maximality and non-minimality, and that that is in fact possible in degree questions. Another example demonstrating this might be (104) in an appropriate context (e.g., an artist wanting to make a realistic life-size sculpture of a polar bear).

- (104) How tall can a polar bear be?

The enforced mention-some interpretation might be what makes (99b,c), odd: a predicate like *know* seems to favor exhaustive interpretations. So in order to interpret (99b,c), one might have to use a slightly disfavored way of combining the question meaning with *know*.

8. TWO APPROACHES TO FLEXIBLE EXHAUSTIVITY

We have reviewed a number of arguments, partly taken from the existing literature, which show that questions do not uniformly receive a (weakly or strongly) exhaustive interpretation. Jointly these arguments undercut an approach in which exhaustivity is built directly into the basic meaning of the question. However, we are also convinced by Groenendijk and Stokhof's arguments that at least in some cases (strong) exhaustivity is required, especially when we are dealing with an embedding verb like *know*.

¹⁴ Note that here also, a maximality operator would give the wrong results: (99b) would come out as equivalent to (99a), while the interpretation of (99c) would be undefined.

EXHAUSTIVITY IN QUESTIONS

We therefore conclude, following Heim (1994), that a flexible approach to exhaustivity is called for, one in which the basic denotation of questions is a non-exhaustive one, but where exhaustivity may arise as a result of several factors that are, so to speak, external to the question itself. We believe that the three formal notions that we have discussed in this paper (the Hamblin/Karttunen denotation, *answer1*, and *answer2*) may play a key role in articulating such an approach. The general approach we advocate immediately raises several important questions as to when and how exhaustivity of either variety comes into play. We do not have a definitive answer to these questions, but in the remainder of this paper we would like to explore two possible ways one may go about answering them. The two approaches have in common that they assign to the question the non-exhaustive Hamblin/Karttunen denotation as its basic interpretation. They differ, however, in the way in which weak and strong exhaustivity come about.

8.1. *Lexical Semantics of Question-Embedding Predicates*

On the first approach, exhaustivity is built into the meaning of certain question-embedding predicates. So for instance, we can distinguish the exhaustive interpretation of the verb *know* from its mention-some interpretation as follows:

- (105) a. $\text{know}_{\text{exhaust}}(w)(x, Q)$ iff $\text{know}_{\text{prop}}(w)(x, \text{answer2}(w)(Q))$
 b. $\text{know}_{\text{mentionsome}}(w)(x, Q)$ iff
 $\exists p[Q(w)(p) \ \& \ \text{know}_{\text{prop}}(w)(x, p) \ \& \ p(w)]$

Here ' $\text{know}_{\text{prop}}$ ' is the denotation of the propositional attitude verb *know* that takes a *that*-complement, and ' $\text{know}_{\text{exhaust}}$ ' and ' $\text{know}_{\text{mentionsome}}$ ' are relations between a person and a question intension (and a possible world) which are defined in terms of ' $\text{know}_{\text{prop}}$ '. Whether we do this with a meaning postulate or by means of lexical decomposition is immaterial for our present purposes. A person x stands in the ' $\text{know}_{\text{exhaust}}$ '-relation to a question-intension Q in a world w iff x stands in the ' $\text{know}_{\text{prop}}$ '-relation to $\text{answer2}(w)(Q)$ in w . Person x bears the ' $\text{know}_{\text{mentionsome}}$ '-relation to Q in w iff in w x stands in the ' $\text{know}_{\text{prop}}$ '-relation to at least one of the propositions in $Q(w)$.

Similarly, we can account for the contrast between the 'transparent' and the 'non-transparent' senses of *write* as follows:

- (106) a. $\text{write}_{\text{transp}}(w)(x, Q)$ iff $\text{write}_{\text{prop}}(w)(x, \text{answer2}(w)(Q))$
 b. $\text{write}_{\text{nontransp}}(w)(x, Q)$ iff $\text{write}_{\text{prop}}(w)(x, \text{answer1}(w)(Q))$

8.2. *Type Shifts*

The second approach treats the operations that turn the Hamblin/Karttunen denotation into either *answer1* or *answer2* as type shifting operations that turn a set of propositions (the Hamblin/Karttunen denotation) into a proposition.¹⁵ Type shifting is triggered whenever there is a mismatch between the type of argument required by the embedding predicate and the basic type of the embedded question. Some predicates, like *wonder* (Groenendijk and Stokhof's intensional verbs), inherently take a complement of the type of a question intension, $\langle s, \langle p, t \rangle \rangle$ (where p is the type of a proposition, $\langle s, t \rangle$). For such verbs, no type shifting is necessary. Other verbs – which are extensional in Groenendijk and Stokhof's sense – take propositional complements, of type p . If their complement is an embedded question, it is necessary to apply a type shift. We can now view *answer1* and *answer2* as type shifting operations which lower an object of type $\langle s, \langle p, t \rangle \rangle$ to one of type p . The answer operations could be operations available in the syntax.

To deal with the mention-some interpretation a third type shift operation can be defined, which shifts the question intension to a generalized quantifier over propositions, that is, an object of type $\langle \langle s, \langle p, t \rangle \rangle, t \rangle$ (where p is the type of propositions $\langle s, t \rangle$):

$$(107) \text{ answer3}(w)(Q) = \lambda P[\exists p[P(w)(p) \ \& \ Q(w)(p) \ \& \ p(w)]]$$

$\text{Answer3}(w)(Q)$ is the set of all those sets of propositions that contain at least one proposition from the Hamblin/Karttunen denotation $Q(w)$ which is true in w . In other words, it is the set of all properties that some true element of the Hamblin/Karttunen denotation has. If a question complement is interpreted on the mention-some reading, its basic question denotation is lifted to the quantifier type $\langle \langle s, \langle p, t \rangle \rangle, t \rangle$. Because of the type mismatch between the matrix verb *know* and the embedded question, the latter has to undergo quantifier raising and be adjoined to the matrix clause, leaving a trace (variable) of type p . (As an alternative one could use Cooper-storage or an argument type shift on the verb, or whatever mechanism one would also use to combine a quantified NP-object with an extensional verb.)

To give a concrete example, the mention-some interpretation of (108) is derived via the (simplified) LF in (109), where the question comple-

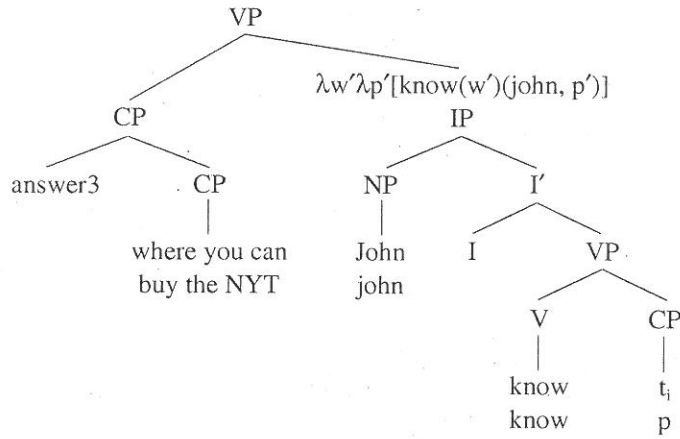
¹⁵ The approach we suggest in this section shares a certain general outlook with the proposals for type shifting among different meanings for questions put forward in Groenendijk and Stokhof (1989). We believe the two approaches are in principle compatible, but since the possible meanings for questions that they consider overlap only partially with the ones we do here, actually combining the two will lead to further complications that we do not want to get into for now.

EXHAUSTIVITY IN QUESTIONS

ment, which has undergone type shifting via answer3, has been quantifier-raised to the root node of the tree. The resulting translation can be reduced as in (110) (where ‘where-you-can-buy-the-NYT’ is an abbreviation for the Hamblin/Karttunen question intension ‘ $\lambda w'\lambda p[\exists x[\text{place}(w')(x) \ \& \ p = \lambda w[\text{you can buy the NYT at } x \text{ in } w]]]$ ’).

(108) John knows where you can buy the New York Times.

(109) answer3(w)(where-you-can-buy-the-NYT')
 $(\lambda w'\lambda p'[\text{know}(w')(\text{john}, p')])$



(110)
 answer3(w)(where-you-can-buy-the-NYT')($\lambda w'\lambda p'[\text{know}(w')(\text{john}, p')]$)
 iff $\lambda P[\exists p[P(w)(p) \ \& \ \text{where-you-can-buy-the-NYT}'(w)(p) \ \& \ p(w)]]$
 $(\lambda w'\lambda p'[\text{know}(w')(\text{john}, p')])$
 iff $\exists p[\text{know}(w)(\text{john}, p) \ \& \ \text{where-you-can-buy-the-NYT}'(w)(p) \ \& \ p(w)]$
 iff $\exists p[\text{know}(w)(\text{john}, p) \ \& \ \lambda w'\lambda q\exists x[\text{place}(w')(x)$
 $q = \lambda w''[\text{you can buy the NYT at } x \text{ in } w'']](w)(p) \ \& \ p(w)]$
 iff $\exists p[\text{know}(w)(\text{john}, p) \ \& \ \exists x[\text{place}(w)(x) \ \&$
 $p = \lambda w''[\text{you can buy the NYT at } x \text{ in } w'']] \ \& \ p(w)]$

This will be true iff John knows at least one true proposition of the form ‘You can buy the New York Times at x’, where x is a place – which is the desired mention-some interpretation.

Of course, to avoid overgeneration, type shifting has to be constrained. In principle, each of the three type shifting operations (answer1, answer2, and answer3) is always available, but we assume that various external factors can be identified to explain why in fact we find only certain specific readings in many examples. At this point, we have no concrete proposals to make as to what these factors might be, however.

The relative pros and cons of the two approaches we have sketched will be clear. In the first approach it is possible to specify exactly for each (extensional) question-embedding predicate what sort of interpretation it gets. But since in a sense this is done by brute force, this approach gives up the hope of achieving a really explanatory account of when we get which reading. The second approach aims to provide just that, but it would be fair to say that at this point this is not much more than a promissory note.

It is possible that the truth is somewhere in the middle, that is, that there is a lexical as well as a grammatical possibility to type-shift. Certain shifts seem pretty much lexicalized (e.g., *know* plus answer2), others seem to apply in a more flexible way. (111) might be a case in point, since *believe* does not normally take an interrogative argument.

(111) You won't believe who I met last night.

Obviously, (111) is interpretable, and gets interpreted using the answer (in some sense) to the question 'who I met last night'. It seems undesirable, though, to allow a lexical type shift from propositions to questions in the case of *believe*: the relative grammaticality of (111) is rather unusual, since *believe* normally does not combine with a question complement. The occurrence of the negation in (111) can be expected to play a role in the explanation of the example's relative well-formedness. Note that this would mean that nonlexical properties of the question-embedding context would have to be taken into account.

Quite generally, if we assume lexical type shift in this case, we can no longer state that *believe* does not select interrogative complements as a rule. Having only grammatical type shift, on the other hand, if we allow it in cases like (111), would predict no differences between verbs that always allow interrogative complements, like *know*, on the one hand, and *believe*, on the other. Thus we think one might want to have both. More will have to be said about when a shift is lexicalized, whether all of them can be lexicalized, and so on.

8.3. *Unembedded Questions*

We have not so far considered the case of unembedded questions and their relation to answers. We will not formally do so in this paper. We believe, rather uncontroversially, that the relation is in essence pragmatic. So when a speaker S asks a question Q, a hearer under most circumstances (though not always) infers that S wants to know a satisfactory answer to Q. What is satisfactory for S depends on the specific context. It may be answer1, answer2, or just an example answer, a true element of Q(w). The hearer will

EXHAUSTIVITY IN QUESTIONS

provide what information s/he can in accordance with Gricean maxims, so in particular the answer will be true and as informative as necessary, but no more than that. If the context suggests that S will be satisfied by an example answer, a hearer will not bore S with a complete list. On the other hand, it seems a natural strategy to provide a maximum of information, that is, answer1. This in turn carries the implicature that the hearer really was as informative as possible; that is, given an answer A it is often inferred that A is the complete answer. S thus concludes answer2 from answer1. If this inference is not desired, the answer provided must be marked as partial (by adding something like *for example, among others, . . .*).

What we have sketched above should extend to degree questions in particular: If for instance John in fact read five books, no well-informed person would answer the question 'How many books did John read?' with 'John read four books', since the answer is, while true, not the most informative one. Giving the most informative answer involves no extra trouble, so it should be very highly preferred. Since the answer actually given carries the implicature that it is the most informative answer, the answer 'John read four books' would even be very misleading.

In the case of unembedded questions, there are various other formal relations possible between question and answer. See, for example, Groenendijk and Stokhof (1984) for discussion. What is important for our purposes is that we believe that the Hamblin/Karttunen denotation will work well for unembedded questions as well as for embedded ones. That is, we believe that the various relations of pragmatically 'good' answers to Q can be defined given the information that the Hamblin/Karttunen denotation Q provides. Although we do not formally show this, we feel justified in this assumption since from the Hamblin/Karttunen denotation the Groenendijk and Stokhof denotation can be derived, and Groenendijk and Stokhof have demonstrated in detail the usefulness of the latter in defining question-answer relations.

9. SUMMARY AND CONCLUSION

We have extended the Hamblin/Karttunen semantic analysis of questions with various notions of what it means to be an answer to a question, following Heim (1994). We have shown that the resulting semantic system can account for exhaustivity and maximality – two effects which the original Hamblin/Karttunen analysis did not capture. In a sense, we have defended a Hamblin/Karttunen semantics against alternatives like Groenendijk and Stokhof (1982, 1984) and Rullmann (1995); however, we have also adopted

important aspects of Groenendijk and Stokhof's and Rullmann's analyses in order to account for their observations, and have thus substantially added to Hamblin's and Karttunen's original proposals.

Concerning degree questions, we have shown that the Heimian notions of answerhood provide a more general explanation for the maximality effect than the maximality operator, since sometimes minimality or a list answer is called for, depending on the semantic properties of the question. Our solution is more flexible and able to account for all types of degree questions.

We agree with Rullmann (1995) that maximality and exhaustivity are really one and the same thing. Accordingly, the maximality effect is captured by what was introduced by Heim to account for exhaustivity: her concept of a complete true answer to a question. Thus we follow Heim in saying that exhaustivity is not a property of the basic question denotation, but comes about through the way question meanings get integrated semantically and interact with their linguistic context.

We thus end up with a fairly rich system of semantic notions concerning interrogatives: the Hamblin/Karttunen denotation, answer1, answer2, and answer3. We believe that there are good arguments that this rich system is in fact useful. There seem to be predicates using answer1 instead of or in addition to answer2, and we think that there are ones which use the Hamblin/Karttunen denotation either directly or indirectly.

By way of presenting arguments for our richer system, we have considered a number of data indicating flexibility in the way an interrogative sentence gets interpreted. The intended interpretation can be marked syntactically in certain ways (*alles, for example, zoal*), but it need not be. It is clear that semantic theories of interrogatives need to capture this flexibility. It is less clear what will be the best way to go about that task.

Our overall strategy is to have a rather weak question denotation, from which other information is recovered by applying certain semantic operations. Starting from a Hamblin/Karttunen denotation, it is possible to strengthen the information contained in it in various ways. This gives one a range of interpretational possibilities, all of which, we argue, play a role in natural language. We hope that external factors can be identified that determine which interpretation is actually chosen. Groenendijk and Stokhof pursue the opposite strategy: They start with a question denotation that contains the maximum of information one can get – information equivalent to answer2. There are thus cases in which they would have to get rid of surplus information, so to speak. We believe that our strategy is more suited to treat natural language interrogatives. It allows for greater flexibility, in what we think is a more natural way. In addition to that, we have

EXHAUSTIVITY IN QUESTIONS

argued that semantic notions might play a role that are not, as far as we can see, recoverable in a Groenendijk and Stokhof system. If our arguments are considered valid, we have a real advantage over a Groenendijk and Stokhof system (or in fact any system that treats strong exhaustivity as a property of the basic semantics of interrogative sentences). In this respect, our proposal can be seen as a defense of a Hamblin/Karttunen semantics for interrogatives.

REFERENCES

- Beck, Sigrid: 1996a, 'Quantified Structures as Barriers for LF Movement', *Natural Language Semantics* 4(1), 1–56.
- Beck, Sigrid: 1996b, *Wh-Constructions and Transparent Logical Form*, PhD dissertation, University of Tübingen.
- Berman, Stephen: 1991, *On the Semantics and Logical Form of Wh-Clauses*, PhD dissertation, University of Massachusetts at Amherst.
- Cresti, Diana: 1995, 'Extraction and Reconstruction', *Natural Language Semantics* 3(1), 79–122.
- Dayal, Veneeta: 1996: *Locality in WH Quantification*, Kluwer, Dordrecht.
- Enç, Mürvet: 1986, 'Towards a Referential Analysis of Temporal Expressions', *Linguistics and Philosophy* 9(4), 405–426.
- Frampton, John: 1990, 'The Fine Structure of Wh-Movement and the Proper Formulation of the ECP', unpublished manuscript, Northeastern University.
- Groenendijk, Jeroen and Martin Stokhof: 1982, 'Semantic Analysis of Wh-Complements', *Linguistics and Philosophy* 5(2), 175–233.
- Groenendijk, Jeroen and Martin Stokhof: 1984, *Studies on the Semantics of Questions and the Pragmatics of Answers*, PhD dissertation, University of Amsterdam.
- Groenendijk, Jeroen and Martin Stokhof: 1989, 'Type-Shifting Rules and the Semantics of Interrogatives', in G. Chierchia, B. H. Partee, and R. Turner (eds.), *Properties, Types and Meaning, Volume II: Semantic Issues*, Kluwer, Dordrecht, pp. 21–68.
- Hamblin, C. L.: 1973, 'Questions in Montague English', *Foundations of Language* 10, 41–53.
- Heim, Irene: 1991, 'Artikel und Definitheit', in A. von Stechow and D. Wunderlich (eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung/Semantics: An International Handbook of Contemporary Research*, de Gruyter, Berlin/New York, pp. 487–535.
- Heim, Irene: 1994, 'Interrogative Semantics and Karttunen's Semantics for *Know*', in R. Buchalla and A. Mittwoch (eds.), *IATL 1*, Akademon, Jerusalem, pp. 128–144.
- Jacobson, Pauline: 1995, 'On the Quantificational Force of English Free Relatives', in E. Bach, E. Jelinek, A. Kratzer, and B. H. Partee (eds.), *Quantification in Natural Languages*, Kluwer, Dordrecht, pp. 451–486.
- Karttunen, Lauri: 1977, 'Syntax and Semantics of Questions', *Linguistics and Philosophy* 1(1), 3–44.
- Lahiri, Utpal: 1991, *Embedded Interrogatives and Predicates that Embed Them*, PhD dissertation, MIT.
- Link, Godehard: 1983, 'The Logical Analysis of Plurals and Mass Terms: A Lattice-Theoretical Approach', in R. Bäuerle, C. Schwarze, and A. von Stechow (eds.), *Meaning, Use and Interpretation of Language*, de Gruyter, Berlin, pp. 303–323.
- McCloskey, James: 1995, 'Quantifier Float and Wh-Movement in Irish English', handout for paper presented at NELS 26, Harvard/MIT, Cambridge, Massachusetts.
- Reis, Marga: 1992, 'The Category of Invariant *alles* in Wh-Clauses', in *Arbeitspapiere des*

SIGRID BECK AND HOTZE RULLMANN

- Sonderforschungsbereichs 340, Bericht Nr. 7*, University of Tübingen, pp. 1–33.
- Rooth, Mats: 1985, *Association with Focus*, PhD dissertation, University of Massachusetts at Amherst.
- Rooth, Mats: 1992, 'A Theory of Focus Interpretation', *Natural Language Semantics* 1(1), 75–116.
- Rullmann, Hotze: 1995, *Maximality in the Semantics of Wh-Constructions*, PhD dissertation, University of Massachusetts at Amherst.
- Rullmann, Hotze and Sigrid Beck: 1998, 'Reconstruction and the Interpretation of Which-Phrases', in G. Katz and S.-S. Kim (eds.), *Proceedings of the Tübingen Workshop on Reconstruction* (Arbeitspapiere des Sonderforschungsbereichs 340), University of Tübingen.
- Rullmann, Hotze and Sigrid Beck: to appear, 'Presupposition Projection and the Interpretation of Which-Questions', in *Proceedings of SALT 8*, Department of Linguistics, Cornell University.

Sigrid Beck
Department of Linguistics, U-145
University of Connecticut
341 Mansfield Road
Storrs, CT 06269-1145
USA
E-mail: sbeck@sp.uconn.edu

Hotze Rullmann
Department of Linguistics, SS 820
University of Calgary
2500 University Drive NW
Calgary, AB
Canada T2N 1N4
E-mail: rullmann@ucalgary.ca