P. Dekker and M. Stokhof (eds.) *Proceedings of the Tenth Amsterdam Colloquium*. ILLC, Univ. of Amsterdam, p. 73-92 (1996).

Degree Questions, Maximal Informativeness, and Exhaustivity*

Sigrid Beck, Universität Tübingen
Hotze Rullmann, Rijksuniversiteit Groningen

1. Introduction

Our aim in this paper is to develop a satisfactory account of degree questions like (1).

(1) How many books did John read?

We focus on the fact that (1) requires an answer naming the maximal number of books John read. In Rullmann (1995) it is argued that maximality in degree questions is one and the same thing as exhaustivity in questions involving individuals. While we adopt this idea as our starting point, we discuss problems for his account with respect to examples like (2), which requires naming the minimal number of eggs that are sufficient.

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

We solve this problem by generalizing Rullmann's (1995) approach. This then leads to a more general discussion of exhaustivity in questions, in which we argue for a flexible approach to exhaustivity. Our paper can be regarded as an extension of a Karttunen (1977) semantics for interrogatives and as a defense of this extended system against various criticisms of Karttunen based on exhaustivity and maximality.
The structure of the paper is as follows. In section 2 we present Rullmann's analysis, which is stated in terms of Karttunen's semantics for questions. Section 3 deals with empirical problems for this analysis. While those problems do not affect the central idea, they do affect the specific implementation in terms of a maximality operator. We will suggest that the various data should receive a uniform explanation in terms of maximal informativeness. To this effect we will adopt in section 4 a proposal by Heim (1994). She reanalyzes exhaustivity not as a property of the semantics of a question itself, but as a property of the notion of answerhood. We will show that (and why) her proposal carries over to degree questions. This leaves us with a satisfactory analysis of degree questions in a Karttunen framework. Section 5 is concerned with the relation between this analysis and Rullmann's (1995) proposal, as well as an analysis along the lines of Groenendijk & Stokhof (henceforth: G&S) (1982, 1984). We will show how in some sense our analysis can remodel their results. However, the differences between our proposal and G&S's semantics are not merely a matter of implementation. Section 6 is devoted to the differences between the two proposals and their potential empirical implications. This includes discussion of

* We are grateful to the audience at the 10th Amsterdam Colloquium and to Irene Heim, Jack Hoeksema, Arnim von Stechow and Wolfgang Sternefeld for comments and discussion. The second author is supported by a grant from the NWO, the Dutch Organization for Scientific Research, as part of the PIONIER-project 'Reflections of Logical Patterns in Language Structure and Language Use'. Research of the first author was supported by the DFG grant to the SFB 340.

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 7 discusses two potential ways of implementing our suggestions formally. We conclude that our proposal ought to be considered as an alternative, flexible approach to exhaustivity, and maximality.

2. Rullmann (1995)

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

More generally, (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', 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).
We will usually 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, the interpretations of (4a) and (4b) 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:

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

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.[1]
In the rest of the paper we will mostly give examples of degree questions involving numbers (*how many*) rather than degrees, for simplicity.

2.3. Individual questions and exhaustivity

Rullmann shows that his analysis can be extended to questions involving individuals rather then 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. $?x\text{: } x=max(\lambda x'[x' \text{ was at the party}])$
d. $\lambda p \exists x[p(w) \,\&\, person(x) \,\&\, p = \lambda w'[x = max(\lambda x'[x' \text{ was at the party in } w'])]]$

[1] 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 7.3.).

Following a suggestion by Jacobson (1995), Rullmann shows that maximality can be used to account for the property of strong exhaustivity argued for by G&S (1982, 1984). G&S distinguish two kinds of exhaustivity in questions, weak and strong exhaustivity. 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.

G&S (1982) propose a semantics for questions which is an alternative to Karttunen's theory but which accounts for both weak and and strong exhaustivity (unlike Karttunen's analysis, which captures only weak exhaustivity).
Rullmann (1995) argues that by introducing maximality into the Karttunen semantics of questions we get a theory that like G&S's theory accounts for both weak and strong exhaustivity. First, 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 G&S 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) $\iota p \exists x[p(w)\ \&\ \text{person}(x)\ \&\ p = \lambda w'[x = \text{max}(\lambda x'[x' \text{ was at the party in } w'])]]$

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:

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

This proposition contains all and only those worlds in which Mary, Sue and Jane were at the party and no one else was. Now if John stands in the know-relation to this proposition this will imply that (i) for every x such that x is a member of {Mary, Sue, Jane}, John knows that x was at the party, and that (ii) 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 G&S 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 G&S'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, John knows that Bill did not read n books, for any n>5. 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 G&S. 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 enough. The interpretation we would get for (13) in Rullmann's (1995) analysis, however, is given in (14):

(14) ?n: n=max(λn'[n' eggs are sufficient to bake this cake])

This is not a satisfactory interpretation for (13), for two (related) reasons. Firstly, 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 sufficient, while we intuitively want the smallest such number, namely three.So 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(λn'[n' eggs are sufficient to bake this cake])

There are a few other predicates that 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 perceivable?

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 dergrees to smaller degrees. So for instance in (17) the question predicate (17b) allows inferences from a number n to numbers m smaller than n; i.e. 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. λn'[John read n' books in w']

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

(18) A predicate P is downward scalar iff
For all n,m: P(n) & m$\leq$n -> P(m)

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

(19) A predicate P is upward scalar iff
For all n,m: P(n) & n$\leq$m -> P(m)

So 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. We do not at present know why this should be so; it explains, however, why they were first overlooked.

We think that the difference between the maximality inducing examples and the minimality inducing ones boils down to informativity. 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 have come to the conclusion 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 informativity 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 with degree questions.

3.2. Degree questions with nonscalar predicates

Consider (20):

(20) a. With how many people can you play this game?
b. How many courses are you allowed to take per semester?
c. How high can a helicopter fly?

A complete answer to (20a) could be, for instance, between 4 and 6. This is, in effect, a complete list of all true answers to the question, or to put it differently, their conjunction. Similarly for the other examples.
The question predicates in (20) are predicates that do not allow inferences either from large degrees to smaller ones or the other way around. If it is permissible to take five courses per semester, for example, then nothing follows about the possibility of taking six courses or four courses. You might be required to take at least five courses. On the other hand, six might be too many. In other words, in cases like (20) we know that there might be a lower bound as well as an upper bound for the degrees that the predicate applies to. More complicated scenarios are conceivable, for example that you are allowed to take either 4 courses or else between 6 and 8. Or a game may be played with any even number of players.
Hence, the question predicates in (20a-c) are neither downward scalar nor upward scalar. We will refer to them as nonscalar 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.

4. Maximal informativeness of answers

We are now in a position to formalize our idea that informativeness is the crucial notion in describing the types of answers you get in degree questions. Our strategy will be to incorporate informativeness not into the semantics of the question, but into the definition of answerhood to a question. We will take as our starting point the ordinary Karttunen semantics for questions. On the basis of that we will define the concept of a maximally informative answer. As it turns out, the notion of maximally informative answer that we need for degree questions, has already been formalized as a concept of answerhood in the Karttunen system by Heim (1994), who calls it answer1.

(21) the answer1 to a question Q in w answer1(Q)(w) = $\cap$(Q(w))

Later on, we will see a second concept of answerhood, answer2. To see how answer1 works, we will now consider an example for each of our three types of question predicate.

The easiest case are the nonscalar predicates. Intersection of all propositions in the Karttunen denotation of the question is just conjunction of all those propositions in which the question predicate is truthfully applied to its argument. So for instance in (22a), given the Karttunen denotation (22b)

(22) a. How many courses are you allowed to take?
b. λp∃n[p(w) & p=λw'[you are allowed in w' to take n courses]]

the intersection of the propositions in (22b) will be the conjunction of all the true propositions of the form "you are allowed to take n courses". So for instance if you are actually allowed to take either four or between six and eight courses, answer1(⟦(22a)⟧)(w) would be the following proposition:

(23) λw[you are allowed to take 4 courses in w and you are allowed to take 6 courses in w and you are allowed to take 7 courses in w and you are allowed to take 8 courses in w]

In the case of a downward scalar question like (24a) answer1(⟦(24a)⟧)(w) would be as in (24b):

(24) a. How many books did John read?
b. ∩(λp∃n[p(w) & p = λw'[John read n books in w']])
c. λw'[John read 5 books in w']

Now suppose that John actually read five books (and not more than five). The proposition that John read four books (which is also in the Karttunen denotation of (24a)) is actually a superset of the proposition that he read five books. Similarly for the other true propositions of the form "John read n books". So the intersection of all these true propositions is the same set as the proposition that John read five books, (24c). answer1(⟦(24a)⟧)(w) is thus identical to the maximum answer.

Finally, consider a minimum case like (25a). Answer1(⟦(25a)⟧)(w) would be constructed as in (25b):

(25) a. How many eggs are sufficient?
b. ∩(λp∃n[p(w) & p= λw'[n eggs are sufficient in w']])
c. λw'[3 eggs are sufficient in w']

Let us assume once more that three eggs are sufficient (and fewer than three eggs are not sufficient). The true propositions in the Karttunen denotation are of the form "n eggs are sufficient" for $n \geq 3$. The proposition that three eggs are sufficient is a subset of all the propositions in that set. Therefore, the intersection

of all the propositons in the set is identical to the proposition that three eggs are sufficient, (25c). We thus end up with the minimum answer as answer1(⟦(25a)⟧)(w).
Note that answer1 can only be empty if the original Karttunen denotation was empty already. Otherwise, it will always contain at least the actual world.
Note also that it can only make a difference whether we take the maximal, minimal or intersective answer if there is more than one proposition in the original Karttunen denotation (i.e. if there is more than one true "simple" answer to the question). Frequently, there is only one proposition in that set, for instance in cases involving modal necessity. Cases with modal possibility are frequently cases in which the Karttunen denotation is not a singleton. That is why so many of our examples involve modals like *can* or *be allowed to*. What type of answer we get then depends on the inferential properties of the predicate.
So for the three types of degree questions that we have looked at, the notion of answer1 seems to give good results. We will now relate this notion to Rullmann's original proposal as well as to G&S's semantics for questions. In the course of doing that we will also get back to the issue of exhaustivity.

5. Strong exhaustivity and answer2

Compare our denotation of answer1 for example (26a), (26b), to Rullmann's (1995) semantics in (26c) (assuming the same facts about the actual world as before):

(26) a. How many books did John read?
b. λw[John read five books in w]
c. λw[max(λn[John read n books in w])=5]

The two propositions are not identical. While (26c) contains the information that five is the maximal number of books John read, (26b) expresses just the proposition that John read five books. Rullmann's semantics and answer1 also differ in (27):

(27) a. Who was at the party?
b. λw[Mary+Sue+Jane were at the party in w]
c. λw[max(λx[x was at the party in w])=Mary+Sue+Jane]

(27c) expresses the proposition that the maximal group that was at the party consists of Mary, Sue and Jane. (27b) 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, (27b) gives the complete true answer, while (27c) gives the complete true answer plus the information that this is the complete true answer to the question.
Rullmann shows that (27c) in a sense captures the same information as the G&S denotation (28) of (27a)(unfortunately, we cannot introduce G&S's theory here, for reasons of space. The reader is referred to G&S (1982)):

(28) λw'[λx[x was at the party in w']=λx[x was at the party in w]]

So essentially, while Rullmann's proposal and G&S's semantics incorporate strong exhaustivity, our notion of answer1 only yields weak exhaustivity.

Discussion of this issue in the last decade has made clear at least that we need to have strong exhaustivity at some points, for example in questions embedded under the verb *know*. Fortunately, Heim's (1994) paper already contains a proposal of how to get strong exhaustivity from answer1, namely by means of her notion of answer2.

(29) answer2(Q)(w) = λw'[answer1(Q)(w') = answer1(Q)(w)]

Heim (1994) shows that answer2 will in general produce the same truth conditions as the G&S-semantics. However, in certain cases this equivalence breaks down. See Heim (1994) for discussion.
This means that we can obtain the information needed to capture strong exhaustivity from the original Karttunen denotation, by applying answer1 and answer2.
Answer2 in a sense remodels a G&S semantics. Moreover, we think that a G&S semantics gives a fairly good result for degree questions, although we don't have the room to show this.[2] This raises the question of why we went through all this trouble of defining notions of answerhood on the basis of a Karttunen semantics, if we could have had a satisfactory result in a G&S semantics straightforwardly without such a fuss.[3] This question will be addressed in the next section.

6. Arguments for a flexible approach to exhaustivity

Summarizing our discussion so far, we have a theory of questions which makes available at least three distinct semantic objects that are associated with a question. Firstly, there is the Karttunen denotation, the set of all true propositions that count as (not necessarily exhaustive) answers to the question. Let's call this set Q(w) (Q being the Karttunen intension). Secondly, we have answer1(Q)(w), the proposition that is the intersection of all the members of Q(w). This constitutes the weakly exhaustive true answer to the question. Thirdly, we have answer2(Q)(w), which is the strongly exhaustive answer to the question and which is (almost) the same as the denotation that G&S assign to questions. 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 G&S have argued so forcefully throughout their work? In this section we will address this question, arguing that having all three notions allows us to adopt a more flexible theory that takes into account cases in which insisting on strong exhaustivity gives rise to truth conditions that appear to be stronger than is supported by our intuitions. In taking this position we have been greatly 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 G&S's. The points we are going to make are not necessarily problematic for G&S when taken

[2]G&S don't treat degree questions explicitly.
[3]In addition to strong exhaustivity, G&S use another argument against Karttunen's theory, namely the distinction between what they call the de re and the de dicto readings of the head noun of the wh-phrase. We do not deal with this important issue in this paper, because it is orthogonal to the problem of exhaustivity. We hope to work out a suggestion in future work.

individually - G&S explicitly discuss and account for some of them, in particular the mention-some interpretations. 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 G&S, 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.

6.1. Embedded questions without strong exhaustivity

As Heim points out, although it is possible to define answer2 in terms of answer1, and answer1 in terms of the Karttunen denotation, this is crucially a one-way street. When we have only answer2 (or in fact any strongly exhaustive question interpretation), it's not possible to get back answer1 or the Karttunen denotation. So in a certain sense, answer2 contains less information than answer1 and the Karttunen denotation. For instance, because a question and its negation impose the same partition on the set of possible worlds, (30a) and (30b) will have the same answer2, but their Karttunen denotation and answer1 will generally differ:[4]

(30) 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. A case in point are emotive factives such as *surprise* (Berman 1991, Heim 1994). (31b) may very well be true although (31a) 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:

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

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 example illustrating the same problem are propositional attitude verbs which refer to ways of conveying information. The verbs we have in mind include

---

[4]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((30a))(w) will be (i), and answer1((30b))(w) will be (ii).
(i) λw[Mary+Sue+Jane were at the party in w]
(ii) λw[Bill+Graham+Marc were not at the party in w]
Answer2((30a),w) will be (iii), and answer2((30b),w) will be (iv):
(iii) λw[answer1((30a))(w) = λw'[Mary+Sue+Jane were at the party in w']]
(iv) λw[answer1((30b))(w) =λw'[Bill+Graham+Marc were not at the party in w']]
However, these two sets of possible worlds will be identical, since whenever the answer1 to (30a) 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.

tell, read, write down and *list*. 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 G&S 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 (32a) and (32b) are not equivalent - and in fact tend to think this is the sense in which this class of verbs is ordinarily understood:

(32) 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 uttering answer1'. The point is that in this sense the verbs would make use of answer1 as well as answer2.

A third case in which answer1 seems to play a role in embedded wh-constructions is based on the semantics of the noun *answer* (Heim (1994)). She notes that (33) 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.

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

So suppose that Mary and Sue were the only party guests then (33) is true if John knows the propositions 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.

6.2. Mention-some readings

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

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

(34a) 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 NYT, but only one particular location, say, the newsstand at the train station. G&S account for the existence of the mention-some interpretation in terms of disjunctions of questions -- an analysis which we don't have the space to 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 Karttunen interpretation of the embedded question. On the mention-some interpretation

(34a) will be true iff John knows at least one of the propositions in the Karttunen denotation of the embedded question. Note that the mention-some interpretation represents a case in which not only strong, but also weak exhaustivity fails.
An important question is when the mention-some interpretation is actually available. For many speakers it only seems to be possible in principle if the question contains a modal element of some sort. We don't know how to explain this apparent influence of modality. Other factors that obviously play a role in the availability of mention-some interpretations are of a pragmatic nature, including considerations of plausibility and world knowledge.

6.3. (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 (35):

(35) 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 (36)), and in German its cognate *so* (see (37)).

(36) Jan wil weten wie er zoal (niet) op het feest waren.
Jan wants know who there zoal (not) at the party were
"John wants to know who for example were (not) at the party"

(37) Hans will wissen, wer so (?nicht) auf dem Fest war.
Hans wants to know who so (not) at the party 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 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. The 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 to appear), and in Dutch we find the corresponding *allemaal*:

(38) 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"

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

To these examples from our own native languages we add (40) from a dialect of Irish English discussed by McCloskey (to appear):

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

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:

(41) Hans weiss wo man alles/ueberall die NYT kaufen kann
Hans knows where one all/everywhere the NYT buy can
"Hans knows where all you can buy the NYT."

In contrast to (34a), (41) 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:

(42) 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."

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

(43) $\text{alles}(Q)(w) = \lambda p[p=\cap(Q(w))]$

Alles operates on a question denotation and gives us the weakly exhaustive interpretation, i.e. the set containing answer1. Since we propose to deal with mention-some interpretations via the elements in the Karttunen denotation, from (43) 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 interpretation as the corresponding question without it. However, we feel that this is not the case: (41) differs in meaning from (34a) in that the former does not allow a nonexhaustive interpretation whereas the latter does. A rigid approach to exhaustivity will have no way to deal with this difference (for instance the G&S approach to the mention-some interpretation could not, in our judgment, predict that (41) does not have a mention-some interpretation, since *alles* could make no difference to the original question interpretation).

6.4. Degree questions with *at least/at most*

The next argument gets us back to the issue of degree questions. Consider the paradigm in (44):

(44) 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 hoechstens da?
How many people were at most there
"How many people were there at most?"

The intuition is clear that (44a-c) mean something different. This holds also for the embedded case:

(45) 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 hoechstens da waren.
Hans knows how many people at most there were
"Hans knows how many people were there at most"

(45b) and (45c) are actually a bit odd. We will come to a tentative explanation for that in a minute.
For (45a) to be true, Hans has to know the exact number of people who where there. For (45b) to be true, he has to know a reasonable lower bound of the number of people who were there, for (45c) 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 (45c).
For the following formal discussion we will assume that *at least* and *at most* mean exactly what they normally do, namely (46)

(46) a. at least n (N) (P) $\Leftrightarrow$ $card(\lambda x[N(x) \,\&\, P(x)]) \geq n$
b. at most n (N) (P) $\Leftrightarrow$ $card(\lambda x[N(x) \,\&\, P(x)]) \leq n$

(47) a-c are the Karttunen denotations of (44) a-c:

(47) a. $\lambda p \exists n[p(w) \,\&\, p=\lambda w'[\exists x[people(x) \,\&\, card(x)=n \,\&\, were\_there(x)(w')]]]$
b. $\lambda p \exists n[p(w) \,\&\, p=\lambda w'[card(\lambda x[people(x) \,\&\, were\_there(x)(w')]) \geq n]]$
c. $\lambda p \exists n[p(w) \,\&\, p=\lambda w'[card(\lambda x[people(x) \,\&\, were\_there(x)(w')]) \leq n]]$

Now suppose that actually 86 people were there. Then (47) a-c are the sets given in (48) a-c:

(48) 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:

(49) a. λw[86 people were there in w]
b. λw[at least 86 people were there in w]
c. λw[at most 86 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 (45b). The ordinary G&S interpretation runs into the same problem.[5]

What is going on here? We think that in (45b,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 (44)b,c and (45)b,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), because 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 (50) in an appropriate context (e.g. an artist wanting to make a realistic life-size sculpture of a polar bear).

(50) How tall can a polar bear be?

The enforced mention-some interpretation might be what makes (45)b,c, odd: A predicate like *know* seems to favour exhaustive interpretations. So in order to interpret (45b,c), one might have to use a slightly disfavoured way of combining the question meaning with *know*.

7. 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 and separately, these arguments undercut an approach in which exhaustivity is built directly into the basic meaning of the question. However, we are also convinced by G&S's arguments that at least in some cases (strong) exhaustivity is called for, especially when we are dealing with an embedding verb like *know*. 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

---

[5]Note that here also, a maximality operator would give the wrong results: (45b) would come out as (45a), while (45c) is undefined.

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 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 Karttunen denotation as its basic interpretation. They differ however in the way in which weak and strong exhaustivity come into play.

7.1. 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:

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

Here $\text{know}_{\text{prop}}$ is the denotation of the propositional attitude verb *know* that takes a *that*-complement.
$\text{know}_{\text{exhaust}}$ 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 answer2(Q)(w) in w. 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" sense of *write* as follows:

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

7.2. Type shifts

The second approach treats the operations that turn the Karttunen denotation into either answer1, answer2, or a mention-some interpretation as type-shifting operations that turn a set of propositions (the Karttunen denotation) into a proposition. 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* (G&S's intensional verbs) inherently take a complement of the type of a question-intension, $<s,<p,t>>$ (where p is the type of a proposition, $<s,t>$). For such verbs, no type-shifting is necessary. Other verbs - which are extensional in G&S'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 $<s,<p,t>>$ to one of type p. To get the mention-some interpretation, a somewhat more complex operation is needed. Intuitively, what we want to do is to existentially quantify over the propositions that are in the extension of the embedded question. Technically one way of implementing this is by turning the question-intension into a generalized quantifier over propositions, of type <<s,<p,t>>,t>. Let's call this operation answer3:

(53) answer3(Q)(w) = $\lambda P[\exists p[P(w)(p) \ \& \ Q(w)(p)]]$

This generalized quantifier can now be combined with the question-embedding verb by the standard techniques that are used to combine a quantified object-NP with an extensional verb (QR, quantifying-in, storage, or type shifts).

(54) know (answer3(Q)(w))(John)(w)
iff $\lambda P \exists p[P(w)(p) \ \& \ Q(w)(p)] \ (\lambda w' \lambda q[know(q)(John)(w')])$
iff $\exists p[know(p)(John)(w) \ \& \ Q(w)(p)]$

In principle, each of the three typeshifting operations is always available, but ideally 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, i.e. 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. (55) might be a case in point, since *believe* does not normally take an interrogative argument.

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

7.3. Unembedded questions

We have not considered so far the case of unembedded questions and their relation to answers. We will not formally relate them 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 all) 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, an element of Q(w). The hearer will 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, i.e. 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, such an answer 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 G&S for discussion. What is important for our purposes is that we believe that the Karttunen denotation will work ok for unembedded questions as well as embedded questions. That is, we believe that the various relations of pragmatically "good" answer to Q can be defined given the information that the Karttunen denotation Q provides. Although we do not formally show this, we feel justified in that assumption since from the Karttunen denotation the G&S denotation (more or less) can be derived, and G&S have demonstrated in detail the usefulness of that in defining question-answer relations.

8. Summary and conclusion

We have suggested a Karttunen semantics for degree questions which captures the maximality effect without using a maximality operator in the question semantics. In our semantics, the maximal answer has no special status in the question denotation at all. This gives us a greater flexibility, which, we claim, is needed. This is demonstrated by degree questions requiring a minimum answer, or ones requiring an explicit list of all true members of the Karttunen question denotation. What type of answer we get is determined by semantic properties of the question predicate: Upward scalar predicates determine minimum answers, downward scalar predicates determine maximum answers, and non-scalar predicates require complete lists. We always get the most informative answer. We therefore link questions to their most informative answers by a notion of answerhood defined in Heim (1994), thus accounting for maximality without assuming a maximality operator, plus minimality, plus list answers. In the context of individual questions, our strategy to this point can account for weak exhaustivity, but not strong exhaustivity. This can be remedied by defining a second notion of answerhood (cf. Heim).
We are thus left with a fairly rich system of semantic notions concerning interrogatives: the Karttunen denotation, answer1 and answer2. 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 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 is 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 to it. Starting from a 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. G&S pursue the opposite strategy: They start with a question denotation that contains the maximum of information one can get (i.e. what we end up with in the "optimal" case). There are thus cases in which they 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 a greater flexibility in what we think is a more natural way. However, we acknowledge that this is in a sense a plausibility argument, and that technically it might be feasible to do things the other way around.

In addition to that, we have argued that semantic notions might play a role that are not, as far as we can see, recoverable in a G&S system. If our arguments are considered valid, we have a real advantage over a G&S 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 Karttunen semantics for interrogatives.

References

Beck, Sigrid (to appear). Quantified Structures as Barriers for LF Movement. Natural Language Semantics

Berman, Stephen (1991). On the Semantics and Logical Form of Wh-Clauses. Ph.D. Dissertation, University of Massachusetts at Amherst.

Groenendijk, Jeroen and Martin Stokhof (1982). Semantic analysis of wh-complements. Linguistics and Philosophy 5: p.175-233.

Groenendijk, Jeroen and Martin Stokhof (1984). Studies on the Semantics of Questions and the Pragmatics of Answers. Dissertation, University of Amsterdam.

Heim, Irene (1994). Interrogative Semantics and Karttunen's Semantics for Know. In Rhonna Buchalla and Anita Mittwoch (eds.) IATL 1. Akademon, Jerusalem.

Jacobson, Pauline (1995). Quantification in Natural Languages. In: Emmon Bach, Eloise Jelinek, Angelika Kratzer and Barbara H. Partee (eds.) Quantification in Natural Languages. Kluwer, Dordrecht.

Karttunen, Lauri (1977). Syntax and Semantics of Questions. Linguistics and Philosophy 1: p. 3-44.

McCloskey, James (1995). Quantifier Float and Wh-Movement in an Irish English. Paper presented at NELS 26, Harvard/MIT.

Reis, Marga (1992). The Category of Invariant *alles* in Wh-Clauses. In: Arbeitspapiere des Sonderforschungsbereichs 340, Bericht Nr. 7, 1-33.

Rullmann, Hotze (1995). Maximality in the Semantics of Wh-Constructions. Ph.D. Dissertation, University of Massachusetts at Amherst.