Exhaustive Interpretation: A Discourse-Semantic Account

Een wetenschappelijke proeve
op het gebied van de Letteren

PROEFSCHRIFT

Ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op
maandag 23 juni 1997,
des namiddags om 1.30 uur precies

door

Rosemarijn Scharten
geboren op 2 januari 1965 te Boxtel
Promotor: Prof. dr. P.A.M. Seuren

Manuscriptcommissie: Prof. dr. M.B.J. Stokhof, voorzitter
                   Prof. dr. S.C. Levinson
                   Prof. dr. F. Zwarts

ISBN 90-9010507-7

© R. Scharten
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Contents</td>
<td>3</td>
</tr>
<tr>
<td>Preface and acknowledgements</td>
<td>5</td>
</tr>
<tr>
<td>1 Linguistic Implicature Theory</td>
<td>6</td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>6</td>
</tr>
<tr>
<td>1.2 A brief history of Radical Pragmatics</td>
<td>7</td>
</tr>
<tr>
<td>1.2.1 Filling the gap between logic and language</td>
<td>7</td>
</tr>
<tr>
<td>1.2.2 Solving logical and linguistic problems</td>
<td>10</td>
</tr>
<tr>
<td>1.3 Exhaustive interpretation in discourse</td>
<td>13</td>
</tr>
<tr>
<td>1.4 Theoretical and methodological considerations</td>
<td>16</td>
</tr>
<tr>
<td>1.4.1 Pragmatic methodology</td>
<td>16</td>
</tr>
<tr>
<td>1.4.2 A theory of discourse structure</td>
<td>16</td>
</tr>
<tr>
<td>1.4.3 Questions as context</td>
<td>18</td>
</tr>
<tr>
<td>1.4.4 Terminological and typographical conventions</td>
<td>20</td>
</tr>
<tr>
<td>2 Radical Pragmatics</td>
<td>22</td>
</tr>
<tr>
<td>2.1 Radical Pragmatics</td>
<td>22</td>
</tr>
<tr>
<td>2.1.1 Why implicature?</td>
<td>22</td>
</tr>
<tr>
<td>2.1.2 Horn scales and more</td>
<td>25</td>
</tr>
<tr>
<td>2.1.3 Expression alternatives</td>
<td>28</td>
</tr>
<tr>
<td>2.1.4 The role of context</td>
<td>29</td>
</tr>
<tr>
<td>2.1.5 The (lexical) semantics of scalar terms</td>
<td>31</td>
</tr>
<tr>
<td>2.1.6 Tests for implicatures</td>
<td>33</td>
</tr>
<tr>
<td>2.1.7 The epistemic modification</td>
<td>36</td>
</tr>
<tr>
<td>2.1.8 Summary</td>
<td>37</td>
</tr>
<tr>
<td>2.2 Theoretical Problems for Radical Pragmatics</td>
<td>38</td>
</tr>
<tr>
<td>2.2.1 Defining scalar implicature</td>
<td>38</td>
</tr>
<tr>
<td>2.2.2 Identifying the expression alternative</td>
<td>41</td>
</tr>
<tr>
<td>2.2.3 Scalar implicatures and truth conditions</td>
<td>44</td>
</tr>
<tr>
<td>2.2.4 Tests for implicature</td>
<td>46</td>
</tr>
<tr>
<td>2.2.4.1 Implicatures and presuppositions</td>
<td>47</td>
</tr>
<tr>
<td>2.2.5 Summary</td>
<td>48</td>
</tr>
<tr>
<td>2.3 Observational Issues</td>
<td>49</td>
</tr>
<tr>
<td>2.3.1 Entailment</td>
<td>49</td>
</tr>
<tr>
<td>2.3.2 Tests for implicature</td>
<td>54</td>
</tr>
<tr>
<td>2.3.3 Inadequacy of predictions</td>
<td>57</td>
</tr>
<tr>
<td>3 Numerals in Context</td>
<td>58</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>58</td>
</tr>
<tr>
<td>3.2 Formal underpinnings</td>
<td>61</td>
</tr>
<tr>
<td>3.2.1 The semantics of specifying be</td>
<td>64</td>
</tr>
<tr>
<td>3.2.2 Relative informativeness</td>
<td>68</td>
</tr>
<tr>
<td>3.3 Exclusion inferences and exhaustive interpretation</td>
<td>71</td>
</tr>
<tr>
<td>3.4 The formal semantics of numerals</td>
<td>73</td>
</tr>
<tr>
<td>3.5 Exhaustive and non-exhaustive interpretation of numerals</td>
<td>75</td>
</tr>
<tr>
<td>3.5.1 Numerals in comment-position</td>
<td>76</td>
</tr>
<tr>
<td>3.5.2 Numerals not in comment-position</td>
<td>80</td>
</tr>
<tr>
<td>3.5.3 Yes/no-questions</td>
<td>85</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3.6 Conclusion</td>
<td>89</td>
</tr>
<tr>
<td>4 Exhaustive Interpretation in Discourse</td>
<td>91</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>91</td>
</tr>
<tr>
<td>4.2 Information structure and the exclusion inference</td>
<td>92</td>
</tr>
<tr>
<td>4.2.1 Lexical material in comment position</td>
<td>92</td>
</tr>
<tr>
<td>4.2.1.1 Mutually exclusive predicates</td>
<td>93</td>
</tr>
<tr>
<td>4.2.1.2 Scalar predicates</td>
<td>96</td>
</tr>
<tr>
<td>4.2.2 Lexical material in non-comment position</td>
<td>99</td>
</tr>
<tr>
<td>4.2.2.1 Mutually exclusive predicates</td>
<td>99</td>
</tr>
<tr>
<td>4.2.2.2 Scalar predicates</td>
<td>100</td>
</tr>
<tr>
<td>4.2.3 Restrictive and nonrestrictive relative clauses</td>
<td>102</td>
</tr>
<tr>
<td>4.3 Exhaustive Interpretation</td>
<td>103</td>
</tr>
<tr>
<td>4.3.1 Two properties of exhaustive interpretation</td>
<td>103</td>
</tr>
<tr>
<td>4.3.2 Escapes from exhaustive interpretation</td>
<td>107</td>
</tr>
<tr>
<td>4.3.2.1 Contradicting or avoiding exhaustivity</td>
<td>107</td>
</tr>
<tr>
<td>4.3.2.2 Objecting to the question: confusion regarding metalinguistic negation</td>
<td>108</td>
</tr>
<tr>
<td>4.3.2.3 Correcting the exhaustivity</td>
<td>110</td>
</tr>
<tr>
<td>4.3.2.4 Non-exhaustive questions</td>
<td>112</td>
</tr>
<tr>
<td>4.3.3 The exclusion inference</td>
<td>113</td>
</tr>
<tr>
<td>4.4 Exhaustive interpretation as a form of Q1 implicature</td>
<td>116</td>
</tr>
<tr>
<td>4.5 Strong vs. weak theorists</td>
<td>118</td>
</tr>
<tr>
<td>4.6 Summary and conclusion</td>
<td>122</td>
</tr>
<tr>
<td>bibliography</td>
<td>126</td>
</tr>
<tr>
<td>samenvatting</td>
<td>130</td>
</tr>
<tr>
<td>curriculum vitae</td>
<td>134</td>
</tr>
</tbody>
</table>
In the spring of 1987, Levinson's *Pragmatics* (1983) was compulsory reading for the course in linguistic pragmatics I was taking. This was my first encounter with the theory of conversational implicatures. And it was Levinson's example *John has fourteen children* and its accompanying explanation (p. 106) that gave me the feeling that something was wrong about the explanation proposed within implicature theory. However, in 1987 I had no idea that this original inspiration would eventually lead to the thesis developed and defended in the pages to follow. The road from this original inspiration to the work that lies before you has been a long one, and not an easy one. There are several people and institutions that I would like to thank for their contributions.

First of all I should thank the former AiO Netwerk "Taal, Logica en Informatie", which has developed into the Onderzoeksschool "Logica", for financing this research. The Faculties of Philosophy and Arts of the Catholic University of Nijmegen provided the necessary facilities. Special thanks go to all at the department of General Linguistics and Dialectology for their hospitality and generous friendship. The pleasant atmosphere at the department has contributed in no small way to my pleasure in work.

There are several people whom I would like to thank for their time and ideas. No doubt they will find the traces of their influence throughout this work; needless to say, the way in which their suggestions were integrated remains entirely my responsibility. First of all I would like to thank Jeroen Groenendijk for his very detailed comments on earlier texts of mine, notably on an early version of chapter 3. Secondly I would like to thank Jan van Kuppevelt for numerous discussions on his and my own work; the synthesis which these discussions led to will be obvious from chapter 1. And finally, but importantly, I would like to thank Steve Levinson. Unknowingly, he provided the initial inspiration for this thesis. Years later, he read large parts of this thesis in an early version; his comments encouraged me greatly to finally finish this work.
1 Linguistic Implicature Theory

1.1 Introduction

This dissertation deals with a set of linguistic phenomena commonly known as implicatures induced by Grice's first submaxim of Quantity. Though the phenomena themselves were not newly discovered, the views developed by Grice's followers in the tradition of Radical Pragmatics are quite widely accepted these days. However, Radical Pragmatics, as I will show in the next subsection, has inherited from Grice some fundamental assumptions about the nature of semantics and pragmatics, assumptions which are not consistent with present-day developments in these area's. Specifically Grice's views on the prime importance of classical predicate calculus as the semantics of natural language have been overtaken by recent developments; and if one, as Grice and his followers do, defines the object of pragmatics negatively as all those phenomena which fall outside the range of semantics, then any change in the views on semantics will have its repercussions on the theory of pragmatics.¹

The theoretical context for this study is that of a particular form of dynamic semantics. There are various forms of dynamic semantics available these days, which have in common that they try to account for the dynamic process of updating a particular context with the information brought in by a new statement made in that context. Apart from this common property, the differences are substantial. One of the prominent differences is that some theories assume, between the level of the semantics of the sentences and that of the world to which they are related, a third level of semantic representation. Specifically, it is not the truth value of the sentences that is assessed, but that of the semantic representation obtained after adding the sentence to the context. Examples of this type of theory are Discourse Representation Theory (Kamp (1981), Kamp and Reyle (1993)) and Discourse Semantics (Seuren (1985)). A form of non-representational dynamic semantics is Dynamic Predicate Logic (Groenendijk and Stokhof (1991)). The semantic theory adopted here is Discourse Semantics, but several aspects of this theory which are used in this study it shares with other forms of dynamic semantics; this is the case notably for the notion of updating a context with a sentence (or accommodating a sentence to a context), which plays an important role in the definitions of relative informativeness given in section 3.2.2.

Section 1.2 contains a brief history of Radical Pragmatics, discussing the philosophical origins of the theory of conversational implicatures and its development into a linguistic theory. In section 1.3 I will present a preview of the claims I will be making in this book. The theoretical context of this dissertation will be sketched in section 1.4.

¹ This fact is acknowledged by other authors as well. Notably Levinson (1988, in prep.) points out that "a rethink of the pragmatics/semantics interface is in order", and one of the reasons for this is formed by the recent developments in semantics.
1.2 A brief history of Radical Pragmatics

1.2.1 Filling the gap between logic and language

Already in the 1960's, it was well known that formal semantics, based on modern predicate calculus, did not capture all of the meaning (informally understood) of natural language. What was not captured by semantics was delegated to the field of pragmatics. One of the most influential theories in the field of pragmatics, both in philosophy and linguistics, has been that of Grice (1967, 1975, 1978, 1989)\(^2\). Grice wanted to contribute to the debate about the alleged “divergences in meaning” between the logical operators \(\neg\), \&, \(\Delta\), \(\Delta\), \(\forall x\), and \(\exists x\), and their natural language counterparts \(not\), \(and\), \(or\), \(if\), \(all\), and \(some\). His claim was

that the common assumption of the contestants that the divergences do in fact exist
is (broadly speaking) a common mistake, and that the mistake arises from
inadequate attention to the nature and importance of the conditions governing
conversation. (24)

In other words, there are no divergences between logic and the semantics of natural language: formal semantics (logic) accounts adequately for the (strict, truth-conditional) meaning of natural language. If an utterance conveys more than its strict, truth-conditional content, then this is caused (among other things) by “conditions governing conversation”. Thus, applied to the logical operators and their natural language counterparts: the semantics of natural language \(and\) is adequately accounted for by the logical operator \(\&\). The fact that sentences with \(and\), such as \(He\ turned\ the\ key\ and\ started\ the\ engine\), sometimes suggest a temporal and/or causal connection between the two events linked by \(and\), is due to factors outside the semantics of \(and\), notably to conditions with respect to the use of language in conversation.

Grice introduced the term \(implicature\) for whatever an utterance conveys beyond its truth-conditional content, for whatever the speaker intends to convey, suggest or imply. The term \(implicature\) thus covers a large set of inferences observed in the use of language, which have in common that they are not logically valid. There are two categories of implicatures, namely the conventional and the nonconventional implicatures. The latter category has two subclasses, the conversational and the nonconversational implicatures. It is the class of nonconventional conversational implicatures which is most clearly developed. These are the implicatures that depend on a specific form of “conditions governing conversation”\(^4\). The starting point for formulating these specific conditions (26-27) is the assumption that conversations are cooperative efforts, in which the participants, at least to some extent,

---

\(^2\) In his (1964), Grice first developed his ideas about non-logical inferences in language. The full theory of implicatures was made public in 1967, when Grice held the William James Lectures at Harvard University. Parts of these lectures were published in later years (1975, 1978), and the full text was published in (1989). Henceforth I will only refer to the latter volume.

\(^3\) Grice includes the symbol \(\forall x\) and its natural language counterpart \(the\) in this list; however, \(\forall x\) is not generally considered to be a logical operator.

\(^4\) In the quote above, Grice (1989:24) suggests that all implicatures are induced by "conditions governing conversation". However, later on in his text (1989:26) he identifies one particular kind of implicatures, namely the nonconventional conversational implicatures, "as being essentially connected with certain general features of discourse." This is probably best understood as implying that all types of implicature are somehow dependent on conditions governing conversation (in a wide sense), and that the subset of conversational implicatures is particularly strongly connected to conditions governing conversation (in a narrow sense).
behave rationally. What counts as rational behaviour is expressed in the Cooperative Principle (CP):

**Cooperative Principle**

Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or directions of the talk exchange in which you are engaged.

Subsequently, Grice formulates four more specific principles guiding rational behaviour in conversation; these, he says, “yield results in accordance with the CP”. Thus, the four maxims of conversation can be seen to be subprinciples of the CP:

**Maxim of Quantity**

1. Make your contribution as informative as is required (for the current purposes of the exchange)
2. Do not make your contribution more informative than is required.

**Maxim of Quality**

Try to make your contribution one that is true, specifically
1. Do not say what you believe to be false.
2. Do not say that for which you lack adequate evidence.

**Maxim of Relation**

Be relevant.

**Maxim of Manner**

Be perspicuous, specifically
1. Avoid obscurity of expression.
2. Avoid ambiguity.
3. Be brief (avoid unnecessary prolixity).
4. Be orderly.

As an illustration of how these maxims are supposed to work, consider the following examples, each exhibiting a particular property of implicatures.

(1)  

A  Do you know where Margaret is?

B  I saw her bicycle outside the supermarket.

Speaker A can only understand B’s utterance as an answer to his question on the assumption that B wants to give information relevant to the ongoing discourse, thus on the assumption that B is adhering to the maxim of Relevance. What B is doing is providing information from which the answer to A’s question can be inferred: from the fact that Margaret’s bicycle is standing outside the supermarket it can be inferred that she probably is at the supermarket. This is the implicature. It is an example of a *particularized* conversational implicature, in which the implicature is determined by specific aspects of the utterance and the context. By changing the context, the implicature can be removed (cancelled):

(1’)  

A  Do you know where Margaret left her bicycle?
I saw it outside the supermarket.

There are also implicatures which are less dependent on specific aspects of the context. Grice calls these generalized conversational implicatures and describes them as follows: “Sometimes one can say that the use of a certain form of words in an utterance would normally (in the absence of special circumstances) carry such-and-such an implicature or type of implicature.” (37) An example is

(2) X is meeting a woman this evening.

By the fact that the speaker has used the indefinite a woman, he conveys (implicates) that the person that X is meeting is “someone other than X’s wife, mother, sister, or perhaps even close platonic friend” (37) In general, Grice suggests, the use of an indefinite NP where a definite one would have been preferred, conveys to the hearer that the speaker is not able or willing to be more specific. This implicature is induced by the first submaxim of Quantity (henceforth Q1), but in a special way: it results from the fact that the speaker did not obey this maxim, or as Grice puts it, flouted this maxim. Presumably, the speaker knows more than he says, and this presumption leads the hearer to infer an implicit message. The first submaxim of Quantity urges the speaker to be as informative as is demanded in a specific situation, and if a speaker does not do this, then this is significant. The maxims can thus be seen to induce implicatures in two different ways: the fact that a speaker obeys the maxims or deliberately and ostensively violates them both give rise to implicatures.

It is clear what Grice means with his notion generalized implicature, but it is also clear that, on the basis of example (2), an entirely different story about this phenomenon can be told. There are some implicit assumptions that Grice makes, that are essential to the induction of the implicit message. Without these assumptions, the implicature would not arise; thus one can argue that it is not the "absence of special circumstances" that makes this particular implicature possible, but that precisely the presence of special circumstances induces the implicature (that it is thus not a good example of the category of generalized implicatures, which it is supposed to illustrate). The crucial point here is the fact that the implicature arises, not its content - what makes a conversational implicature a generalized one is not its content but the conditions under which it arises. In order to clarify the discussion of this example it is necessary to put it in a context (for reasons that will become apparent later on in this chapter, this context will have the form of a question):

(3) A Whom is X meeting this evening?
    B X is meeting a woman this evening.

Now the first assumption that Grice implicitly makes, and that is necessary for the implicature to arise, is the fact that X is a man. If X were a woman, as in (3’),

(3’) A Whom is Ms. Johnson meeting this evening?
    B She is meeting a woman this evening.

the use of a woman would not induce any implicature about the possible identity of this second person, let alone an implicature that something indecent is going on. It does induce another implicature, namely that B does not know the exact identity of the woman Ms. Johnson will be meeting this evening. This reveals the second assumption which is necessary
for the implicature in (2) to go through: B has at least some knowledge of the relevant facts. If B only knows that the person that X is meeting this evening is a woman, then the utterance of (3B) does not carry any associations with indecent activities going on. Suppose that B is a police officer investigating a crime. He is shadowing suspect X and taps his telephone. One day he overhears a brief conversation which makes it clear that X has an appointment with a woman, but which does not reveal the identity of this woman. In a situation like this, if B reports to a superior as in (3), the only implicature conveyed is again that B does not know the exact identity of the woman. From these brief remarks we may conclude that Q1 implicatures are very strongly dependent on the knowledge that a speaker possesses and on what the hearer assumes the speaker to know.

Grice’s theory has a very strong intuitive appeal. However, on closer inspection many questions, both of an empirical and of a theoretical nature, arise. A central problem is the lack of clear definitions of fundamental concepts such as implicature and maxim. Anyone who adopts Grice’s theory will inherit these problems, and will thus have to find a solution for them. Another aspect of Grice’s theory which is inevitably inherited by any follower is the importance of logic as a semantics of natural language. The field of pragmatics is formed in a negative way: every semantic aspect of an utterance that does not fit into a logical explanation belongs to the field of pragmatics. The same goes for the term implicature: it covers all those inferences in language that are not valid within classical predicate calculus.

1.2.2 Solving logical and linguistic problems

Grice’s theory gave rise to extensive discussion and follow up, both in philosophy and linguistics. In linguistic pragmatics, the dominant theory is that developed by Horn (1972, 1989). He adopts Grice’s theory of implicature in order to solve a complex of problems in logic and linguistics, and in doing that he follows the same line of argument as Grice does. Grice argues that the nature of the problem with respect to the semantics of natural language connectives is misunderstood: the alleged divergences in meaning between logic and language are not semantic, but pragmatic in nature. There are three problems that Horn wants to solve this way. First is the fact that negation often means “less than”. This problem was identified by Jespersen (1924, 1968:325/6), also quoted in Horn (1972: 32):

Here the general rule in all (or most) languages is that not means “less than”, or in other words “between the term qualified and nothing”. Thus not good means “inferior”, but does not comprise “excellent”; not lukewarm indicates a lower temperature than lukewarm, something between lukewarm and icy, not something between lukewarm and hot. This is especially obvious if we consider the ordinary meaning of negated numerals: He does not read three books in a year/ the hill is not two hundred feet high/ his income is not £200 a year/ he does not see her once in a week/ the bottle is not half full – all these expressions mean less than three, etc.

The second problem is that of what are sometimes called the one- and two-sided readings of a large number of lexical items (the so-called “scalar terms”), including the numerals, various quantifiers, modal operators and gradable adjectives. All these lexical items can receive either of two readings, which can informally be characterized as an “at least” and an

5 These terms are adopted from Burton-Roberts (1984).
“exactly” reading. The linguistic problem is of course whether these items are ambiguous or not. Remarkably, this problem originates not from linguistics but from philosophy, namely from Aristotle’s modal logic. In his De Interpretatione, Aristotle at first confuses two interpretations of the modal operator possible, informally, possible can be taken to mean "possible and maybe necessary" (the one-sided, "at least" interpretation) or "possible but not necessary" (the two-sided, "exactly" interpretation). Confusing these two interpretations leads to a contradiction (see Kneale and Kneale (1962:84)). However, in later work he carefully separates the two interpretations, in the way indicated in table (1).

<table>
<thead>
<tr>
<th></th>
<th>possible-1</th>
<th>possible-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>understood as</td>
<td>not impossible</td>
<td>neither impossible nor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>necessary</td>
</tr>
<tr>
<td>our terminology</td>
<td>one-sided</td>
<td>two-sided</td>
</tr>
<tr>
<td>Kneale and Kneale (1962)</td>
<td>“it is not impossible that”</td>
<td>“it is contingent that”</td>
</tr>
</tbody>
</table>

In their discussion of the distinction between these two readings, Kneale and Kneale suggest that metaphysical arguments are responsible for the distinction between possibility (possible-1) and contingency (possible-2).

Aristotle was probably determined by metaphysical considerations to make contingency, rather than possibility, the leading notion in his theory of problematic syllogisms; for in his metaphysics the distinction between the necessary and the impossible on the one hand and the merely factual on the other is of fundamental importance. (85)

It is thus clear that the question of whether the scalar terms are ambiguous or not, is the linguistic version of a problem which is in origin of a metaphysical nature.

The third problem which Horn intends to solve by adoption of the theory of implicatures is of a logical origin. It concerns the "logic of subcontrariety" (see his 1989), the question of the logical relationship between the subcontraries in Aristotle's square. The subcontraries are Some men are mortal and Not all men are mortal/ Some men are not mortal. In tracing the history of this problem, Horn mentions "the standard practice of relegating this inference to an extralogical domain" (211). He identifies the relationship between the subcontraries as one of conversational implicature: the uttering of some men are mortal induces the Q1 implicature not all men are mortal.

Horn integrates the first two problems in the following quote (1972:31-32), in which he states the problems he intends to tackle in his dissertation:

    b. John has $200.

---

6 The discussion to follow is based on Kneale and Kneale (1962:81-96).
7 My use of the term interpretation at this point should not be taken to imply any position with respect to the linguistic problem of the ambiguity of possible; my aim here is to discuss briefly some of the more philosophical, metaphysical matters involved.
c. John doesn’t have $175.

If 175 is taken in “exactly n” reading, (a) is inconsistent with the state of the world described by (b), i.e. is false if the latter is true; if it is taken in the “at least n” reading, the two are consistent. The negation of the (a) sentence, (c), is normally understood as negating the “at least” reading, so that this negation is inconsistent with (b). If the cardinal number is stressed, however, the negation in (c) can be taken as external, in which case the “exactly n” reading is possible, if not preferred. The external reading of the negation in (c) is, of course, perfectly consistent with (b).

In his (1989: 205-208), we find a different formulation of the same problems, in terms of the fundamental (semantic, truth-conditional) intuitions which ask for an explanation. In the contexts in (5), which of the utterances in (4) would be true and which would not?

(4) a. a is good.
    b. a is not good.

(5) C₁ a is in fact excellent.
    C₂ a is in fact good but not excellent

Under certain conditions [4a] conveys that context C₂ is known to hold, while under other conditions it conveys that either C₂ or C₁ holds. Under certain conditions [4b] denies that either C₂ or C₁ holds, while under other conditions [4b] - or a longer variant of it - simply denies that C₂ holds.

In (5), C₁ represents the one-sided, "at least" interpretation of good, C₂ represents the two-sided, "exactly" interpretation. Sentence (4b) is added to study the interpretation of good under negation. Horn sketches several positions one could adopt with respect to the question asked here: which of the two sentences is true in which of the two situations? His own position is that “[4a] is true (or true but misleading) in C₁, and [4b] false.”

With these complexes of problems in mind, the following line of thought seems to have directed Horn towards Grice’s notion of implicature: in logic, negation affects only the truth-conditional content of an item. Therefore, the aspects of meaning which are not affected by negation are the result of some non-truth-conditional inference, possibly implicature. Now the meaning (informal) of the scalar terms can be analysed as existing of two components: a lower bound, informally rendered as “at least”, thus “at least beautiful”, “at least three”, and an upper bound, rendered as “at most”. With all scalar terms, only the lower bound ("at least") is cancelled by negation, leaving the upper bound intact. This results in the characteristic "less than" or "at most"-interpretation of these items under negation:

(6) a. John does not have three children.
    b. Mary is not pretty.

Thus, the lower bound (“at least”) can be seen to form the truth-conditional meaning of these scalar terms. The upper bound, on the other hand, is presumed to be an instance of non-truth-conditional implicature, induced by Grice’s first submaxim of Quantity (Q1). The Gricean
inference involved is this: the speaker said *Mary is pretty*, where *pretty* (truth-conditionally) means “at least pretty”. Presuming the speaker is obeying the CP, then, by Q1, the speaker has provided the maximum amount of information possible. Thus, he intended to convey that Mary is pretty, but not more than that. In other words, *pretty* is not only the lower bound (“at least”) but also the upper bound (“at most”). The upper and lower bounds together result in the interpretation “exactly”, e.g. “exactly three”, “exactly possible” (i.e., possible and not necessary). The lower bound is a form of generalized conversational implicature, which means that it is present unless it is cancelled, either explicitly or implicitly.

The adoption of the non-truth-conditional notion of implicature simultaneously solves both problems mentioned above, that of the semantics of negation, and that of the apparent ambiguity of scalar terms. To start with the latter, what appear to be two different meanings of these items are in fact an interpretation with and one without the implicature. Scalar predicates, among which are the numerals, are assumed to

assert lower-boundedness - "at least n" - and given tokens of utterances may, depending on context, implicate upper-boundedness - "at most n" - so that the number may be interpreted as denoting an exact quantity.

(1972: 33; see also 1989: 243)

This explains Horn’s judgement on sentence (4a) in context C1 : (3a) is true, because *good* means (asserts) “at least good”, but misleading, because the implicatum “at most good” is not true in C1. This view on the semantics of scalar terms also solves the second problem, the fact that negation sometimes seems to mean “less than”. This is identified not as a fact about the semantics of negation, but about the semantics and pragmatics of scalar terms: negation only denies the truth-conditional lower bound of a predicate’s meaning, leaving the upper bound intact, thus *not three* gets interpreted as “at most three”, i.e. less than three.

Thus, Horn’s solution for these problems is exactly in line with Grice’s work. Instead of reverting to unattractive semantic solutions (the scalar terms are ambiguous between an “at least” and an “at most” reading, or they are vague or semantically general), part of the problem is identified as being pragmatic in nature. Moreover, by the wholesale adoption of Grice’s notion of implicature, Horn implicitly assumes the views with respect to the nature of semantics and pragmatics that are inherent to this work.

1.3 Exhaustive interpretation in discourse

After surveying the Radical Pragmatics literature in chapter 2, I will in chapters 3, 4 and 5 present an alternative analysis for Q1 implicatures. However, in order to be able to develop an alternative analysis, a terminological matter has to be solved. Although the type of implicit inference which is the subject of this study is generally known as Q1 implicature, it is necessary to give it a different name here. The term *Q1 implicature* not only identifies the inference, but simultaneously indicates its origin as a Gricean implicature. It is necessary to choose a term which does not have any theoretical associations. On an intuitive level, the implicit inference concerned can be described as one that excludes, within a certain relevant domain, everything (all possibilities/all possible situations) that is not explicitly mentioned. Thus, in the following dialogue between two children,

(7) Q Who is coming to pick you up this afternoon?
A My mother is coming to pick me up this afternoon.

the answer implicitly conveys (implicates) that no one other than the child's mother is coming to pick him up, thus e.g. not his father, neighbour or grandmother. This intuitive characterization I will summarize in the term exclusion inference. Adequately characterizing this type of inference is the main goal of this study, and in doing that many of the fundamental assumptions of Radical Pragmatics will be challenged.

In chapters 3, 4 and 5 I will argue that the exclusion inference consists of two separate processes, the first of which is purely linguistic in nature and the second of which should be viewed against the background of the structure of human cognition in general. The first process is the process of exhaustive interpretation of an utterance in discourse, the second is the derivation of the exclusion inference (in a narrow sense). In arguing this claim I will not attempt to solve the whole complex of problems that induced Horn to adopt the notion of implicature: the one- and two-sided readings of predicates, the "less than" interpretation of negation and the relation between the subcontraries in Aristotle's square. Only the first of these will receive an alternative explanation.

All recent explanations for the problem of the one- and two-sided interpretations of predicates start from the basic assumption that this is a problem of lexical semantics or pragmatics. That is, the solution to the problem should be found at the level of the semantics or pragmatics of each individual lexical item. Thus, there have been proposals that the lexical items involved are ambiguous (a proposal that Horn argues against very strongly (1972)), or that the two different interpretations are the result of a form of loose speaking or vagueness (see e.g. Sadock (1984), Atlas (1989)), or that we have a unitary meaning with an implicature superimposed. In challenging the latter of these theories, I will argue that the fundamental assumption that the problem of one- and two-sided readings is a problem of lexical semantics and pragmatics is misguided. The difference between the one-sided and the two-sided interpretations of a predicate is the result of processes at the level of discourse. The factor determining the one- and two-sided interpretations of a predicate is the information structure of the utterance in which the predicate at issue appears.

The information structure of an utterance is laid out in terms of topic and comment. The terms topic and comment are notorious for the confusion that exists over their exact meaning and denotation; I will not stray into this discussion. In section 1.4 I will give a precise specification and motivation of my use of these terms, but for introductory purposes we may say that the terms topic and comment denote the information asked for (the question) and the specification given (the answer) in a discourse, respectively.

The relation between the information structure and the possible interpretations of a predicate is this: when a predicate is in comment position, i.e. when it is in a strict sense the answer to a question, it invariably receives the two-sided, "exactly" interpretation. I will say that it is interpreted exhaustively as the answer to the question. When that same predicate is part of the topic expression of an utterance it receives the one-sided, "at least" interpretation. Now the exclusion inference, which in Radical Pragmatics is seen as an inference dependent on the content of a predicate, can be seen as an inference related to the information structure of a discourse. When a predicate a in comment position is interpreted exhaustively, this means that with some emphasis it is asserted that only a holds and nothing else. In other words, all other possibilities are discarded. This can be seen very clearly from the semantics and pragmatics of the numerals, which is the topic of chapter 3. The relation between context and the interpretation of other scalar predicates is the topic of chapter 4.
In these chapters, the process of exhaustive interpretation will play an important role. Intuitively speaking, exhaustive interpretation can be seen to lead to the following type of inference: when an utterance talks of e.g. John stealing pens, then on an exhaustive interpretation the hearer will assume that only John, and no one else, stole the pens. A more adequate characterization requires the presence of a context:

(8) Q Who stole all the pens from your classroom?
A John did.

The exhaustive character of the answer in (8) can be paraphrased in the form of a cleft: *It was John (and no one else) who stole all the pens from my classroom.* It is characteristic of WH-questions that their answers are by default interpreted exhaustively. One can characterize this process in terms borrowed from the judiciary system and say that exhaustivity has to do with "speaking the whole truth", not with "speaking the truth" simpliciter. Although theories of semantics mostly attempt to understand the latter of these, it cannot be denied that in daily life the former is a real and important notion. It surfaces very clearly and explicitly in a court of law, as the aim at such assemblies is to investigate some event: what happened and what did not happen? This investigation takes the form of assessing the truth or falsity of reports on these events. However, the task of "speaking the whole truth" is not limited to a court of law, but exists in less formal situations as well.

One of the crucial questions induced by the line of argument presented here is: which are these "other possibilities" that are discarded by the process of exhaustive interpretation? This problem arises here as much as it does in Radical Pragmatics, where scales are introduced as a partial answer to this problem: a scale is a partially ordered set of semantic concepts such that, if some concept relatively low on the scale is used in discourse, the concepts higher up on that scale are negated. In chapter 2 I will address the problems that this theory runs into, in section 4.3.3 I will make some remarks on the nature of the problem of the "other possibilities" in the context of the theory presented in chapters 3 and 4; no attempt at solving this problem will be made.

In this study, I will not be claiming that implicature in the sense of Grice (i.e., as a type of inference induced by general principles of conversation) does not exist. Grice incorporates too many phenomena under the heading of implicature for it to be possible to make such a general claim on the basis of the study conducted here. The point is that the linguistic version of implicature theory, i.e. Radical Pragmatics, is misguided with respect to the Q1 (notably the scalar) implicatures. One of the phenomena it intends to explain by adopting implicature theory, namely the one- and two-sided readings of a large number of lexical items, is better explained as a purely linguistic phenomenon. For anyone familiar with the Radical Pragmatics literature, many of the examples in chapters 3 and 4 will not be new; what is new however is the view developed on these examples. One very essential aspect of this view is the fact that the context of an utterance inducing an exclusion inference is systematically taken into account, in a way that will be explained in the next section.
1.4 Theoretical and methodological considerations

1.4.1 Pragmatic methodology

Any attempt to write a book on pragmatics, i.e. a study on the use of language, runs into an important methodological problem: one has to decide what the characteristics of language in use are, and how these characteristics can be represented in a theoretical study (in a written form). In Radical Pragmatics, this problem was not addressed; most examples in Radical Pragmatics literature are sentences (in isolation), which is rather a strange methodology for a pragmatic theory. A minimal requirement for doing pragmatics properly is looking at utterances, which are particular instances of sentences. To put it in Austin’s terminology, a sentence is a type, i.e. an object that is adequately characterized as a syntactically well-formed string of words. The object of pragmatics however are tokens, i.e. individual instances of sentences or sentences in use (the usual term for these is utterance). Unfortunately, the methods for representing this object in writing are limited.

The most important difference between a sentence and an utterance lies in the fact that an utterance functions in a context. This is a generally recognized fact, just as it is generally recognized that the exclusion inference is an extremely context-sensitive type of inference. In this book, I will be systematically taking context into account when studying the exclusion inference. There are many aspects to a context which influence the interpretation of an utterance: context is both linguistic (what was said or suggested before) and extra-linguistic (the speech situation, the participants in the discourse). Within the limits of this study, it is not possible to take all of these aspects into account. Therefore, the more limited method of representing the immediate linguistic context of an utterance in the form of a question is adopted. This provides a clear notion of context, which makes systematic study of the interaction between the context of an utterance and the exclusion inference it induces possible.

This dissertation is not intended as a defence of a general strategy of adopting questions as context in pragmatics. In this respect, its pretensions are as limited as indicated above: in the context of a study of a particular type of implicit inference, commonly known as Q1 implicature, representing the context as a question has some interesting consequences. It reveals a kind of interaction between context and inference which was not signalled before, and suggests an alternative explanation for these inferences. In section 1.4.2, the theoretical background to the choice of representing the context in the form of a question is discussed; in section 1.4.3 some further arguments motivating this choice are given.

1.4.2 A theory of discourse structure

The method of representing the context in the form of a question is suggested by the theory of Van Kuppevelt (1991, 1995). This theory, which I will henceforth refer to as Discourse Topic Theory (DTT), deals with the structuring or segmentation of information in discourse. The central notions topic and comment are defined in terms of questions and answers. A discourse is considered to be a hierarchically ordered set of questions and answers. The process of discourse production is a process of contextually inducing explicit or implicit questions, which are subsequently answered. The answers are explicit, they are so to speak the visible

---

8 At this point I take the terms question and answer to refer to linguistic objects, in a largely intuitive sense. No theorizing about them is presupposed. In chapter 3, the syntactic theory supporting the notions of question and answer, and topic and comment, will be discussed in detail.
part of the discourse. The questions are mostly left implicit, not only in monologues but also often in dialogues. Their role in structuring a discourse is assumed to be as follows:

Every contextually induced explicit or implicit (sub)question $Q_p$ that is answered in discourse constitutes a (sub)topic $T_p$. $T_p$ is that which is being questioned; a set of singular or plural (possibly non-existent) discourse entities (or a set of ordered $n$-tuples of such entities in the case of a $n$-fold question) from which one is selected as an answer to $Q_p$. Comment $C_p$ is provided by this answer and names or specifies the entity asked for. (1995:113)

An example will illustrate these notions:

(9) $Q_1$ Who was arrested yesterday?
A$_1$ Tom was arrested yesterday.

In this small discourse we have a question inquiring after the (as yet undetermined) person that was arrested yesterday. $Q_1$ constitutes topic $T_1$, namely the (undetermined) set of persons who were arrested yesterday. Answer $A_1$ provides the determination asked for and specifies one person that was arrested yesterday; comment $C_1$ is thus Tom. Note that topic and comment are both defined at the referential level, topic is the undetermined set referred to by the question, and comment is “that which is asked for”, a subset of the topic-set. However, my main interest in this book will not be with topics and comments as sets of referential objects, but with the linguistic expression of topics and comments, specifically the difference in interpretation between linguistic material that is in comment position versus linguistic material that is in non-comment position.

This definition of the information structure of a discourse in terms of questions and answers allows one to identify topic and comment not only at the level of single utterances, but also at the level of larger discourses. In general, a discourse consists of sequences of question-answer pairs and is considered closed when all subquestions that arose in relation to that topic are answered satisfactorily.

Implicit in DTT is the assumption that every utterance takes up one of three structural positions in a discourse: it is a feeder, an implicit or explicit question or an answer. Feeders can be both linguistic and non-linguistic in nature, and are mainly characterized by their function, which is the initiation or re-initiation of a process of questioning in discourse (either when the context is empty or when all previous discourse topics have been closed off) (see 1991:124). A linguistic feeder is defined as “a topicless unit of discourse or one of which the topic is no longer prominent at the moment of questioning” (1995:7). However, the possibility must be left open for utterances which are "topic-neutral" but do not function as feeders. Examples of this would be public signs such as "don't walk on the grass" or "no pets allowed in this building". Feeders and topic-neutral utterances will appear only sporadically in this study.

The two processes described here, initiating a topic by means of a feeder ($f$), and closing off a discourse topic after all subtopics have been answered satisfactorily, are illustrated here:

(10) $f$ A Tomorrow is John's birthday.
$Q_1$ B What would be a good present for him?
A1  A  Tickets to the Vermeer-exhibition in The Hague would be a good present for him.

Q2  B  Why would that be a good present for him?

A2  A  Yesterday he told me that he would like to go there.

In this discourse, Q2 constitutes a subtopic to the main topic, which is constituted in Q1. A's initial answer, that tickets to the Vermeer-exhibition in The Hague would be a good present for John for his birthday, apparently is not satisfactory for B. Q2 is posed in order to amplify the answer A1 in such a way, that it becomes a satisfactory answer to Q1.

In this section I have presented some aspects of DTT, in so far as they will be relevant to the investigation conducted in the later chapters of this book. Notably the assumption that every discourse is considered to be a hierarchically ordered sequence of (implicit) questions and answers will play a very important role. The questions define topics (of different levels: discourse topics and (sub)topics), their answers provide the comment. One limitation that DTT adopts is the fact that it concerns itself (almost) exclusively with topics and comments as defined by WH-questions. Throughout this book, I will adopt the same limitation, with the exception of a brief discussion of yes/no-questions in section 3.5.3. The term question will be understood to refer only to WH-questions.

1.4.3 Questions as context

There are three reasons for adopting DTT in the context of the study of implicatures. DTT in effect claims that all utterances in a discourse that are not feeders or context-neutral utterances are the answer to a question; this accounts for the internal coherence and the structuring of information in discourse. Because of this unitary view on the immediate linguistic context of an utterance, a systematic study of the interaction between utterance and context becomes possible. The first reason for adopting this theory in the context of a study of implicatures was already mentioned in section 1.4.1: formulating a prior question to which a given sentence can be considered to be an answer is a way of contextualizing that sentence and thus viewing it in the light of a possible context of use. This operation should therefore be of interest to pragmatic theory. The second reason lies in the fact that in most early writings on implicatures, thus in Grice (1967), but also in O’Hair (1969) and more recently in Hirschberg (1991), the majority of the examples consist of a question-answer sequence, where the implicature is induced by (some parts of) the answer. In other words, within implicature theory it is not uncommon or unusual to present the context in the form of a question. Remarkably however, this choice is never motivated.

The final and most important reason lies in the fundamental notion of Q1 implicature. Q1 implicatures are by definition dependent on what could be described as the required level of informativeness at a certain moment in the discourse. This follows directly from Grice’s formulation of the first submaxim of Quantity, “Make your contribution as informative as is required (by the current purposes of the exchange)”. If an inference is dependent on the amount of information that is required at a certain moment, then a definition of this inference presupposes a definition of what this required level of informativeness is. Finding the latter definition is one of the recurring themes in the literature on implicatures, and so far with little hope of success:
(...) specifying what the ‘required’ or ‘appropriate’ level of such informativeness is in some [context] seems impossible to do in any general way. (Hirschberg 1991:13)

The method of contextualizing an utterance adopted here does not provide a general theoretical definition either, but it does provide an escape route. Without having to solve all theoretical problems involved, this problem can be formulated in terms of questions and answers: when every (non-feeder) utterance is the answer to an (implicit or explicit) question, supplying the required level of informativeness amounts to answering the question satisfactorily.

Obviously, there is a theoretical problem here. Presumably, the notion “answering a question satisfactorily” is not easier to define in any general way than “supplying the required level of informativeness”, so one could object that in this respect the approach proposed here does not offer any improvement over the theory of Radical Pragmatics. However, apart from this theoretical problem, using questions as context does provide a clear practical advantage: it locates exactly the point in discourse where informativeness is important. It is not in general possible to determine “the required level of informativeness” at some point in a discourse, but it is often possible to determine what answer would satisfy a certain question asked. When there is no sound theoretical foundation to notions like these, the fact that one is intuitively more graspable than the other is a substantial argument for preferring the one over the other.

One more practical problem of the method adopted here needs to be considered. In a question-answer pair, the question only represents the immediate linguistic context of the utterance that functions as its answer. However, the generation of exclusion inferences is influenced by many more aspects of the wider context, both of the larger linguistic context and of the extra-linguistic context. The extra-linguistic context comprises both aspects of the physical speech situation and extra-linguistic (encyclopaedic or world) knowledge that the speaker and hearer possess. A consequence of this is that the same question, in different contexts, may call for a different answer, e.g.

(11) Two people meet each other for the first time. Both are students at an evening-school which recruits students from all over the country.
   A Where do you live?
   B In Nijmegen.

(12) The same two people, two months later. They have agreed to work on a paper together, and are making an appointment for the following week.
   A Where do you live?
   B At 17, Oranjesingel.

Thus it may seem that the notion "satisfactory answer to a question" is not capable of supplanting the notion "required level of informativeness". However, one may argue that in either situation, a different question was asked. In (11), the question A is synonymous with in what city do you live? or where do you come from?, while in (12), A is synonymous with what is your address? Disambiguation in this way (i.e., by incorporating more information from the wider context into the question) is possible in many situations. Obviously, the method of representing context by way of a question has its limitations, and these have to be taken into account.
1.4.4 Terminological and typographical conventions

The notions topic and comment will play an important role in the chapters 3 and 4, but not in their common function as indicators of information structure. The chief interest will be not in the information structure itself, but in the influence that information structure has on the interpretation of particular linguistic items and on implicit inferences induced by those particular linguistic items. Thus, we will be looking at the relationship between the position that linguistic items occupy within the information structure of the discourse and their interpretation.

Within the theoretical context of DTT, a particular linguistic item α can occupy the following positions: it can be part of a feeder or a topic-neutral utterance, part of the topic-expression of an utterance or part of the comment-expression of an utterance. Only the latter two positions will be considered in this study. For no matter how important feeders are in the daily use of language, our aim is to study the interaction between information structure and interpretation; and for those feeders which are "topicless units of discourse" it is not possible to determine their information structure. Feeders will thus be left out entirely, and for our present purposes a predicate α can only be part of the topic-expression or of the comment-expression of an utterance.

If α is part of the topic-expression of an utterance, then it is also part of the question, due to the definition of topic as a set of possible extensional values constituted by the question. On the other hand, if α is part of the comment-expression of an utterance, it will always be part of the answer. In a strict sense, the comment constitutes the answer to a question. α may exhaust the answer, as answers may consist of a single word, or α may be only part of the answer. The comment-expression of examples will in this study be signalled by underlining it:

(13) Q Where did Peter find the stolen book?
    A Peter found the stolen book in David's room.

The syntactic analysis of questions and answers developed in chapter 3 will allow us to determine precisely, for each question-answer pair, what position in the information structure specific lexical items occupy.

As said before, the object of pragmatics is sentences in use, commonly referred to as utterances. In this book, the information structure will be regarded as one of the identifying characteristics of an utterance. The immediate linguistic context, which has the form of an (explicit or implicit) question, imposes an information structure in terms of topic and comment on its answer. In (14) and (15) we thus have two different utterances, because they function as answer to two different questions9; consequently, their information structure differs:

(14) Q Who ate all of the cookies this morning?
    A John ate all of the cookies this morning

(15) Q When did John eat all of the cookies?
    A John ate all of the cookies this morning.

---

The majority of examples in the Radical Pragmatics literature consists of single sentences, without any context. When studying them in the perspective sketched here it is thus necessary to supply a context in the form of a question. Whenever a context is supplied, this is signalled by putting the question in angled brackets:

(16) <Q How many children does John have?>
A He has fourteen, if not fifteen children.

Thus, any question in angled brackets is not in the original.

The search for the characteristics of the exclusion inference will proceed in the following steps. Chapter 2 contains a detailed thematic discussion of the theory of Radical Pragmatics. In chapter 3, the first steps to an alternative explanation of the exclusion inference will be taken. First, the syntactic and semantic analysis supporting the notions of question and answer will be discussed. Next the interpretation of the cardinal numerals in context will be studied, and the notion of exhaustive interpretation will be introduced. In chapter 4, other scalar predicates and their interpretation in context will be studied, and the notion of exhaustive interpretation will be developed more generally. Thus, chapters 3 and 4 contain a discussion of the first process involved in generating an exclusion inference, namely that of exhaustive interpretation in discourse. The second process, which I have called the derivation of the exclusion inference in a narrow sense, and which involves the determination of which are the relevant alternatives, will be touched upon briefly in section 4.3.3.
2 Radical Pragmatics

2.1 Radical Pragmatics

The discussion of Radical Pragmatics in this chapter will be based chiefly on a limited number of sources, which represent it most clearly. These sources are the work of Horn (notably 1972, 1989), Levinson (1983) and Atlas and Levinson (1981), the formal development of Radical Pragmatics in Gazdar (1979) and the critical development by Hirschberg (1991). In chapter 1 I discussed the problems that induced Horn to adopt the notion of implicature, namely the one- and two-sided readings of the so-called scalar terms, together with their interpretation under negation, and what he calls “the logic of subcontrariety”. The introduction of scalar implicatures, which are a subclass of the Q1 implicatures, was meant to solve these three problems simultaneously. Due to the orientation of Radical Pragmatics, this chapter will deal mainly with scalar implicatures\(^{10}\), although there are no principled differences between the scalar and non-scalar types of Q1 implicature. In this section, I will give an exposition of the theory of Radical Pragmatics; most critical notes will be saved until later sections. Section 2.2 contains a discussion of theoretical problems, section 2.3 a discussion of observational issues.

2.1.1 Why implicature?

As was pointed out in Chapter 1, Horn adopted the Gricean notion of implicatures to solve a complex of problems. In this subsection I will show how Horn's notion of scalar implicatures fits into Grice's scheme. Grice attributed several properties to implicata, not all of which are capable of functioning as criteria. Only three allow for testing: calculability, non-detachability and defeasibility (cf. Sadow 1978).

First of all, implicatures are cancellable or defeasible: from the fact that an implicature is a non-truth-conditional inference it follows that it should be possible to cancel it without a contradiction arising. This may be done either explicitly or implicitly (contextually). Secondly, implicatures are nondetachable:

> Insofar as the calculation that a particular conversational implicature is present requires, besides contextual and background information, only a knowledge of what has been said (or of the conventional commitment of the utterance), and insofar as the manner of expression plays no role in the calculation, it will not be possible to find another way of saying the same thing, which simply lacks the implicature in question, except where some special feature of the substituted version is itself relevant to the determination of an implicature (in virtue of one of the maxims of Manner).

(Grice 1989:39)

\(^{10}\) As the subject of this chapter is the implicature theory of Radical Pragmatics, I will adopt its terminology as much as possible. Consequently, the term exclusion inference, as introduced in the previous chapter, will not be used in this discussion. It will return in chapter 3.
The application of this criterion is severely limited; when there is no other (in many ways equivalent) way of expressing the same content, it cannot apply. For this reason it does not apply in the majority of scalar cases. Finally, implicatures are calculable: a so-called “Gricean argument” can be constructed which shows how the utterance of a certain sentence, on the assumption that the speaker is adhering to the Cooperative Principle, leads to some implicature.

According to these criteria, and in so far as they apply, Horn’s scalar inferences turn out to be genuine implicatures. The prevailing opinion is, first, that cancellation is no problem for any of the categories that Horn includes under the scalar implicatures:

(1)  
   a. John has three children, in fact four.  
   b. Her essay is good, in fact excellent.  
   c. Mary is not just pretty, she's beautiful!  
   d. Not only is it possible she'll win - it's certain she will.  
   e. She's patriotic or quixotic - in fact, she's both.  
   f. You ate some of the cookies - indeed, you ate all of them.

The final criterion of calculability applies as well:

Let us take as an instance my assertion of the sentence in [2a]: why does this statement normally convey the proposition in [2b] if this proposition is not part of its meaning? And why does it seem to exclude [2c]?

[2]  
   a. Pat has three children.  
   b. Pat has exactly three children.  
   c. Pat has four children.

The argument from [2a] to [2b] proceeds as follows:

[3]  
   i. Cardinals like 3 are lower-bounded by their literal or conventional meaning; hence [2a] means (is true iff) Pat has at least three children.  
   ii. There is a stronger statement than [2a], that is, [2c], such that the latter unilaterally entails the former but not vice versa. (Actually, there are in fact infinitely many such stronger statements.)  
   iii. Given Q, if I know or believe that Pat has (at least) four children, and that it would be relevant to you to know this fact, it would be misleading for me to tell you that he has three children.

---

11 Hirschberg (1991) observes that nondetachability can only be seen to hold by a purely theoretic argument. As Horn specifies that his scalar predicates are “not lexical items but rather semantic concepts”, and as nondetachability may be seen as a condition demanding semantic identity, “it would seem that implicatures conveyed via scalar predication must be nondetachable.” (55)

12 Horn distinguishes between cancelling and suspending an implicature; both processes will be discussed in subsection 2.1.6.

13 Q is Horn’s version of Grice’s second submaxim of Quantity. In (1984), Horn proposes a reduction of Grice’s maxims to two basic principles, Q and R:

(i) The Q Principle (Hearer-based):  
MAKE YOUR CONTRIBUTION SUFFICIENT (Cf. Quantity-1)
iv. You are prepared to assume that I am abiding by the Cooperative Principle and its component maxims, including Q; I know this, as well, and you know I know it. Thus, you take me to be observing Q unless I indicate otherwise.

v. Therefore, you infer that the reason I chose not to express the stronger proposition in [2c] is that I didn’t know for a fact that it was true.

[Schematically, \( \neg K(p) \)]

vi. You infer that for all I know [2c] is false, that is, that Pat has fewer than four children. [Schematically, \( K \neg (p) \)]

vii. If, in addition, you assume that I know how many children Pat has, you can infer that I know (and am informing you) that Pat has fewer than four children; hence (given (i)) that [2b] is in fact the case, that Pat has three children.

(1989: 214)

As a Gricean argument, this is perfectly in order. However, its necessarily informal phrasing poses problems. Ever since Grice’s theory appeared on the market there have been discussions on the exact phrasing of arguments like these (e.g., is step (v) really necessary? Should steps (vi) and (vii) be distinguished, as Horn claims, or be integrated, as Gazdar and Levinson maintain?). These debates have never proved very fruitful. Rather, they are a consequence of the lack of clear definitions, as will be discussed in section 2.2.

The fact that implicatures are not presuppositions or entailments is supported by various arguments. First of all, implicatures are not entailments, because entailments cannot be cancelled; e.g.,

(4) a. I managed to get to the party.
   b. I got to the party.
   c. *I managed to, and in fact didn’t, get to the party.

In (4), (b) is an entailment of (a). An attempt at cancellation, as in (c), leads to a contradiction. A second argument why implicatures are not entailments is that equating them would lead to a reductio ad absurdum (see Gazdar (1979:51), Fogelin (1967:18)):

(5) a. Some of the boys were at the party.
   b. Not all of the boys were at the party.
   c. Some, in fact all, of the boys were at the party.
   d. All of the boys were at the party.

In standard Radical Pragmatics, the uttering of (5a) implicates (5b). (5c) is an example of implicature cancellation, (5d) is a sentence which entails (a) but is inconsistent with the implicature in (b). Now if (b) were an entailment of (a), then, by transitivity, (d) would entail

---

SAY AS MUCH AS YOU CAN (given R)
Lower-bounding principle, inducing upper-bounding implicata

(ii) The R-Principle
MAKE YOUR CONTRIBUTION NECESSARY (Cf. Relation, Quantity-2, Manner)
SAY NO MORE THAN YOU MUST (given Q)
Upper-bounding principle, inducing lower-bounding implicata.
(b) as well. But (b) and (d) are mutually inconsistent. Therefore, it is concluded, (b) cannot be an entailment of (a).

Next, implicatures are not presuppositions because “unlike presuppositions, implicatures do not usually survive if the [implicature-inducing utterance] is embedded in a sentence that does not entail it.” (Gazdar (1979:51) In (6), the implicature induced by the embedded phrase \textit{some of the boys were at the party} in isolation, namely “some of the boys were not at the party”, does not survive the embedding:

(6) Mary managed to tell John when she thought that some of the boys were at the party.

Levinson notes a second argument why implicatures are not presuppositions: contrary to implicatures, presuppositions do “seem to be detachable” (1983:116). That is, for implicatures it is possible to express a certain content in two or more different ways, while all expressions carry the same implicature. His example is one of irony:

(7) John's a mental prodigy.
    John's an exceptionally clever human being.
    John's an enormous intellect.
    John's a big brain.

But presuppositions are typically attached to certain lexical items; consequently, if that particular item is not used, the presupposition is not present either. In (8), (a) presupposes (b), but (c), which according to Levinson is truth-conditionally equivalent to (a), does not carry the presupposition.

(8) a. John didn't manage to reach the summit.
    b. John tried to reach the summit.
    c. John didn't reach the summit.

A third criterion for distinguishing presuppositions from implicatures will be presented in section 2.2.1.4.

2.1.2 Horn scales and more

A Horn scale\textsuperscript{14} (or simply \textit{scale}) is a partially ordered set of semantic concepts. These concepts are ordered by relative informativeness, which is defined as semantic entailment of utterances containing some expression of these concepts:

\[ P \rightarrow Q \iff Q \text{ is true under every assignment of truth values (i.e., in every possible world) under which } P \text{ is true.} \]

(Horn 1972:9)

The members of a scale are ordered from strong to weak (from left to right) by semantic entailment as follows: if \( p_i \) and \( p_j \) are members of the same scale, then if \( Q(p_i) \) entails \( Q(p_j) \) (where \( Q(p_i) \) and \( Q(p_j) \) are identical expressions except for the scalar items), then \( p_i \) is

\textsuperscript{14} The term \textit{Horn scale} was introduced by Atlas and Levinson (1981:33).
considered to be semantically stronger or more informative than \( p_j \). This is conventionally written as

\[(9) \ <p_b, p_j>\]

However, the ordering of scalar concepts cannot be established before the members of the scale are identified, and this is one of the notorious problems of implicature theory. Horn (1989:232) limits himself to giving examples, which illustrate clearly the intuitive nature of a scale:

\[(10)\]

\[
\begin{array}{ll}
<\text{all, most, many, some}> & <\text{always, usually, often, sometimes}>\\
<\text{and, or}> & <\ldots, 6, 5, 4, 3, 2, 1>\\n<\text{must, should, may}> & <\text{necessary, (logically) possible}>\\n<\text{certain, \{probable/likely\}, possible}> & <\text{obligatory, permitted}>\\n<\text{boiling, hot, warm}> & <\text{freezing, cold, cool, (lukewarm)}>\\n<\text{beautiful, pretty, attractive}> & <\text{hideous, ugly, unattractive, plain}>\\n<\text{adore, love, like}> & <\text{loathe, hate, dislike}>\\n<\text{excellent, good, OK}> & <\{\text{terrible/awful}, \text{bad, mediocre}\}>\\
\end{array}
\]

When the scale is determined, it is possible to describe a scalar implicature as an implicature induced by the use of a scalar predicate, and dependent on Grice’s first submaxim of Quantity. The use of one of the weaker elements of a scale, e.g. \textit{That girl is attractive}, in the presence of scalar alternatives such as \textit{pretty} and \textit{beautiful}, induces the inference that the speaker has not chosen the most informative (the strongest) expression available to him. On the assumption that the speaker is observing the Cooperative Principle, there must be a reason why he has not been maximally informative. This reason must be that the speaker is not convinced of the truth of the stronger expressions. By the use of a weak scalar predicate when a stronger one is available, the speaker in effect conveys (intentionally, intendedly) that the stronger ones do not hold: the girl in question cannot be described as pretty or beautiful, as far as the speaker is concerned. Hirschberg (1991:48) rightfully points out that Horn makes a strong claim here: the use of a weaker element from a scale leads to the inference that the stronger elements do not hold (would be false if used), not that their value is unknown.

Although intuitively clear, the notion of a Horn scale has proven hard to define. For that reason, several modifications and elaborations of the notion were proposed within the theory of Radical Pragmatics.

A substantial widening of the notions of a scale and scalar implicature is offered by Hirschberg (1991). In her view, scalar implicatures are supported by partially ordered sets (POSETS)\(^{15}\), which are conceptual orderings. Following such an ordering, partial orderings

---

\(^{15}\) The term \textit{set} is to be understood in a mathematical sense. The orderings that support scalar implicature are \textit{relations}: "Let \( V \) be a set. Then a \textit{RELATION} \( O \) on \( V \) is a subset of \( V \times V \). The ordered pair \( <p_i, p_j> \) \( O \) is commonly denoted by \( p_iO p_j \). The metric by which such a relation is constructed I will term the \textit{ORDERING METRIC} of the relation, or simply, its \textit{METRIC}." (122) Not all mathematically possible relations support scalar implicature, the following types are excluded: "CYCLIC relations, temporal parallelism, and in general, relations that are not both \textit{REFLEXIVE} and \textit{ANTISYMMETRIC} or relations that are not both \textit{IRREFLEXIVE} and \textit{ASYMMETRIC} – as well as relations that are not transitive." (122)
over expressions can be defined. Hirschberg's posets include the Horn scales as a subset. The posets are more liberal than the Horn scales in several respects.

First, while Horn scales are all linear orderings, Hirschberg incorporates hierarchical orderings as well. Examples of these are whole/part, type/subtype, entity/attribute, set/subset and set/member. This is an example involving a set/member relationship:

(11) A. Do you have apple juice?
   B. I have grape or tomato or bloody mary mix.

(... in [11] B evokes the set of juices to implicate ¬BEL(B, I have apple juice); by committing herself to the set \{grape, tomato, bloody mary mix\}, B does not commit herself to the set \{grape, tomato, bloody mary mix, apple\}, and thus, does not commit herself to apple juice. (108/9)

Secondly, Horn limits scales to those orderings which have one-sided entailment as the ordering relation; Hirschberg allows other ordering relations ("metrics") as well, as illustrated in (11). Other examples of posets and their ordering metrics are given in (12).

(12)

\begin{tabular}{|l|l|}
\hline
poset & metric \\
\hline
tort/ misdemeanor/ felony/ capital crime & is a lesser offense than \\
cold/ cool/ tepid/ warm/ hot & is cooler than \textsuperscript{17} \\
hot/ warm/ tepic/ cool/ cold & is warmer than \\
subtype/ type, e.g. rubber cement/ pasta vinaigrette/ salad dressing turtle/pet & is a kind of \\
\hline
\end{tabular}

Thirdly, according to Hirschberg, scalar implicatures involve not only higher or lower values on the same scale, but alternate values as well.\textsuperscript{18} Alternate values are values that share a common subordinate or superordinate; again, (11) is an example of this. The conceptual set involved in (11) can be graphically represented as follows:

---

\textsuperscript{16} In (1989), Horn acknowledges the necessity of allowing other metrics than just one-sided entailment. However, he does not provide any suggestions as to what these might be. He suggests that there is a continuum of types of scales, ordered by their type of metric. On this continuum, the quantitative scales, defined by one-sided entailment, form one end.

\textsuperscript{17} Introducing this metric for this scale solves a problem Horn is faced with: if one-sided entailment is the ordering relationship for a scale, then warm and cold can never be part of the same scale. Horn solves this problem by splitting this, and similar scales, into a positive and a negative part: <lukewarm, warm, hot> and <cool, cold, freezing>. Hirschberg finds this a counter-intuitive solution, which moreover is hard to apply to scales concerning modifiers such as <bad, all right, good>. (1991:116/7)

\textsuperscript{18} The definition of higher, lower and alternate values is: "Relations defined by ordering the non-null members of the power set of some set x by set-inclusion allow a poset representation of x and its non-null proper subsets as follows: Any non-null proper subset of a set may be ranked as LOWER than the set which includes it, and that set, in consequence, will represent a HIGHER value in the ordering. Subsets which are neither included in, not include, one another, will be ALTERNATE values in this poset." (128/9)
The final aspect in which Hirschberg's view on scalar implicatures differs from that of Horn concerns the types of commitment a speaker can convey. In Horn's view, the affirmation or negation of some value on a scale induces a scalar implicature, which is itself an affirmation or negation of other values on the same scale. Hirschberg allows a third type of commitment, namely ignorance. This too works in two ways: if a speaker declares his ignorance with respect to some value, then this induces a scalar implicature:

(13) A. Is it warm in Antarctica in the summer?  
B. I don't know if it gets above freezing.

if the ordering freezing/cold/.../warm/... [is] salient, then B may be seen as declaring ignorance of the lowest value s/he truthfully can. In [2], B implicates that, not only is he unaware of whether it is warm in Antarctica in the summer – but he doesn't even know if it gets above freezing.

(63/4)

And secondly, the implicature itself may be a statement of ignorance, as in (13).

2.1.3 Expression alternatives

The formulation of the Gricean argument in section 2.1.1 exhibits very clearly one of the recurring themes in the theory of Q1 (scalar) implicature. Although the formulations of these arguments differ in many ways, the condition formulated in (ii) above is always present in some form. Condition (ii) states that the generation of a Q1-implicature (not only the scalar ones!) is dependent on the presence or availability of some expression, which “might be desirable as a contribution to the current purposes of the exchange” (Levinson 1983:135). I will refer to this available expression as the expression alternative. This term is adopted from Gazdar (1979:57), who defines it as follows:

Sentences $\phi_A$ and $\phi_B$ are **expression alternatives** with respect to $\alpha$ and $\beta$ iff $\phi_A$ is identical to $\phi_B$ except that in ONE place where $\phi_A$ has $\alpha$, $\phi_B$ has $\beta$.

The difference between Q1 implicatures in general and the subset of scalar implicatures can now be formulated in terms of expression alternatives: it is a general problem for all Q1 implicatures to determine the expression alternatives upon which the implicature is dependent. Scales can be regarded as a partial answer to this problem. They are only a partial answer because of the problems of definition associated with semantic scales, briefly mentioned above. When one regards scales to be semantic in nature, as Horn did in (1972), the problem of defining them is placed outside pragmatics.

The notion of expression alternative also allows us to express clearly the difference between the positions of Horn and Hirschberg (discussed above). While for Horn (1972), with
his semantic, quantitative notion of scale, there has to be a relation of one-sided entailment between the expression alternative and the expression actually used (the expression alternative entails the expression actually used but not vice-versa), Hirschberg lets go of this criterion. She introduces the general term *ordering metric* for the relations that order some set of concepts, upon which the implicature-inducing utterances are based. Moreover, she lets go of the condition that there be some ordering relation between an expression used and the expression alternative in order for there to be an implicature: alternate values may induce scalar implicatures as well.

### 2.1.4 The role of context

Context plays a large role in the generation of scalar implicatures, because they are generalized implicatures (see section 1.1.1). To put it in modern terms, scalar implicatures are default inferences: they arise unless they are cancelled, either explicitly or implicitly (contextually). However, in Horn's work the exact role of context is not worked out. Phrases like “a (...) application of Quantity in a context where the maxim can be appropriately invoked” (1989: 251) or equivalent, occur frequently but do not receive an adequate follow-up. Several other authors however have remedied this omission. I will briefly discuss Gazdar's and Hirschberg's proposals with respect to the role of context in the generation and cancellation of implicatures.

Gazdar (1979) proposes a theory of several semantic and pragmatic aspects of language, which he integrates into one system. Apart from two subtypes of $Q1$ implicature, namely scalar and clausal implicature, he discusses presupposition and logical form. Gazdar defines context simply as unordered sets of propositions “constrained only by consistency” (130). Integration of an utterance (i.e., its logical form, its implicatures and its presuppositions) into the context proceeds in two steps. First, for every utterance, its potential implicatures (clausal and scalar) and potential presuppositions are calculated. Next it is determined which of these become the actual implicatures and presuppositions of that utterance. The process responsible for this latter step is that of “satisfiable incrementation”, which amounts to checking for consistency. The definition of satisfiable incrementation takes as input two sets of propositions, $X$ and $Y$. As output it gives the incrementation of $X$ with $Y$ (graphically, $(X\upharpoonright!Y)$), but only with “those members of $Y$ that can be added to $X$ without ANY risk of an inconsistency arising”. When applied to e.g. potential implicatures, $X$ represents the context (i.e., an unordered set of propositions) and $Y$ the set of potential implicatures. $(X\upharpoonright!Y)$ then represents the new context, resulting from the incrementation of only those implicatures which do not lead to a contradiction (i.e., the actual implicatures) to the old context. Gazdar observes that this process “can be usefully glossed as a formal rendition of the slogan “all the news that fits”.” (1979:131)

This notion of SATISFIABLE INCREMENTATION allows us to capture formally the context-sensitivity of both implicature and presupposition. ONLY those [potential] implicatures and [potential] presuppositions which are satisfiable in the context of utterance actually emerge as the implicatures and presuppositions of the utterance. (132)

---

19 Clausal implicatures will be discussed in section 2.2.2.
Thus, in Gazdar's theory, the notion of consistency with a context, which is a set of propositions, can be seen to function as a filter, which removes (among other things) certain instances of potential implicature. Gazdar's system adopts a type of process which was typical for linguistic theories in those days, and which is entirely in line with Horn's views on generalized implicature: generate liberally, and then remove, by means of some filter, parts of what was generated. This type of process, though it may lead to a correct output, is generally abandoned because of its computational redundancy (why generate structures that never surface?), and therefore its psychological implausibility. A final aspect to note about this view of context is that Gazdar assumes context to be an unordered set of propositions. As was pointed out in chapter 1, and this is not a new observation, linguistic context should not be viewed as an unordered, but as a partially ordered set of propositions.

An entirely different role is attributed to context in Hirschberg's theory. Context does not work as a filter, at the end of the generation process, but as a factor indicating which parts of an utterance may be involved in the generation of scalar implicature. For Hirschberg, the generation of an implicature is dependent on the fact that there is an ordering $O$, such that it is mutually believed salient by speaker and hearer. Context specifies which parts of an utterance are salient and which are not (and thus, indirectly, which orderings are salient and which are not).

In process-terms, Hirschberg's view of the role of context is different both from Gazdar's and Horn's views. In her theory, various different aspects of (both linguistic and non-linguistic) context may indicate that some part of an utterance (some expression) is salient. If speaker and hearer assign the same foci of salience in their respective mental representations of the context, then the conditions are fulfilled for successfully inducing and interpreting an implicature. In other words, the context provides the necessary preconditions for the occurrence of implicature. In this system, there is no need for a process of contextual cancellation such as Horn and Gazdar adopt. Thus, from a computational point of view, this theory is more efficient and consequently more attractive.

However, Hirschberg does not develop an actual theory of context. In chapter 6 of her dissertation Hirschberg reviews work on several factors, linguistic and more general psychological, which might indicate or determine salience of an expression or a relation. Among the linguistic factors she discusses are syntactic, intonational, semantic and pragmatic cues. These include theories of focus, speech acts, planning and goals, and also Rosch's theory of prototypes, which is a purely psychological theory about the organization structure of human cognition.

Summarizing, although context is generally admitted to play a very important role in the generation and/or cancellation of implicature, theories about this role are not always well-developed. Context can be seen to function in several different ways in the different forms of Radical Pragmatics: either as a filter cancelling unwanted instances of implicature (Horn, Gazdar) or as a mechanism determining which are the implicature-inducing parts of an utterance (Hirschberg). However, Gazdar's theory of context is unattractive for reasons of psychological implausibility, and Hirschberg does not develop an actual theory of context.

---

20 Obviously, this is a version of condition (ii) from the Gricean argument in section 2.1.1.
21 Salience is defined as "that which is most likely to be attended to". (1991:65)
22 In her discussion of various theories of discourse and information structure, Hirschberg notes that they all have one aspect in common: the set of potentially salient expressions according to these theories is always limited to NPs and VPs. However, she notes that any item of a sentence may be focussed, notably by contrastive stress. In effect, Hirschberg claims that intonational cues are dominant over semantic and pragmatic cues.
2.1.5 The (lexical) semantics of scalar terms

The theory of Radical Pragmatics not only is a theory of (scalar) Q1 implicatures, but also involves a very specific view on the lexical semantics of scalar terms. Levinson (1983:134) puts it very explicitly:

this theory in effect embodies a claim that the semantic content of lower items on a scale\(^{23}\) is compatible with the truth of higher items obtaining, and the inference that higher items do not in fact obtain is merely an implicature.

So in the case of quantifiers for example,

the meaning of rightmost scalar items like *some* is consistent with leftmost items like *all* in such scales. It follows that *some* has the semantic content paraphrasable as “at least some” or “some if not all”. (144)

Horn (1989:243) expresses the same position:

(... the weaker the scalar item, the wider the conditions under which its containing proposition is true, as exemplified in [14]:

<table>
<thead>
<tr>
<th></th>
<th>True (it is hot)</th>
<th>True (it is very warm)</th>
<th>True (it is warm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>it's boiling</td>
<td>True (it is hot)</td>
<td>True (it is very warm)</td>
<td>True (it is warm)</td>
</tr>
<tr>
<td>it's hot</td>
<td>True (it is hot)</td>
<td>True (it is very warm)</td>
<td>True (it is warm)</td>
</tr>
<tr>
<td>it's very warm</td>
<td>True (it is very warm)</td>
<td>True (it is very warm)</td>
<td>True (it is warm)</td>
</tr>
<tr>
<td>it's warm</td>
<td>True (it is very warm)</td>
<td>True (it is very warm)</td>
<td>True (it is warm)</td>
</tr>
</tbody>
</table>

Thus, in Radical Pragmatics, the semantics of e.g. *some* is "at least some". Or, in Horn's words: *some* "asserts lower boundedness". Horn argues for this position by pointing out that an alternative analysis has unwanted consequences:

If *some* really meant, rather than implicating, ‘some only’, *some at least* would be a contradiction. Furthermore, the radical bilateralist\(^{24}\) would predict that expressions like *some but not all*, and indeed *some only* or *only some* (along with the examples of [15]) would be semantically redundant, which they clearly are not.

[15] reinforcement

Pat has exactly three children

three and only three children

\(^{23}\) Note that Levinson regards scales as ordered sets of lexical items, not as sets of concepts. This becomes clear also from his remarks that adopting Gricean implicature allows for simplifications in the semantics of a great number of lexical items.

\(^{24}\) Horn describes two positions that one can take up with respect to the semantics of scalar terms: the Radical Pragmatics view, in which *some* means "at least some" and the inference to "only some, some but not all" is a scalar implicature, is called the unilateralist position; the opposing view, in which *some* means "some", is called the radical bilateralist position. This opposition will be discussed in greater detail in section 4.5.
This view on the semantics of scalar terms invites two observations. First, the choice one makes with respect to the semantics of the implicature-inducing terms is decisive for one's position in pragmatics. The "at least" semantics for scalar terms is not, as Levinson suggests, a consequence of the adoption of a certain pragmatic theory, it is the other way around: a particular choice of semantic theory determines what problems get relegated to pragmatics. Thus a different lexical semantics for scalar terms will lead to a different pragmatics. And secondly, in the absence of any clear criteria for what is a scalar term, this view leads to absurdity when the scope of Q1 implicature is widened. Consider Grice’s example (1989:37)

(16) John is meeting a woman tonight.

which is said to induce the Q1 implicature "the woman is not John's wife, mother or (platonic) friend. In this example, $P$mother/wife) entails $P$woman). But it cannot possibly be that woman means "at least woman". A similar point can be made in connection with Kempson’s examples (1986:81/2):

(17) She didn't lose a finger: she lost an arm.
(18) I didn't invite John to supper: I invited John, Mary and Susanna

In (17), the implicature-inducing expression is losing a finger. But that does not mean "losing at least a finger; losing a finger and possibly something else"; and similarly for (18): it is absurd to claim that a name $X$ means "at least $X$". In general, there is no way within Radical Pragmatics to avoid this line of argument, to avoid the general view of the semantics of the scalar terms being widened to other lexical items, as there are no criteria for what counts as a scalar term. Moreover, the argument that is applied to scalar items applies to wider classes of cases as well. But obviously it is absurd to claim that the meaning of mutually exclusive items involves some "at least" aspect.

The first doubt over this position with respect to the semantics of scalar items is cast by Hirschberg (1991:116). In discussing Horn's reasons for dividing the scales into a positive and a negative pole25 she writes:

> It is easy to say that beautiful does NOT entail hideous or ugly – but does it entail plain – or even pretty for that matter?

> If we accept that a hot soup is NOT therefore also a cold soup, can we really accept that it IS a warm soup? Whatever hot entails about warm, e.g., A hot soup is not only warm, it's hot., it should also entail about cold. Or are all these values just mutually exclusive states.

Most Radical Pragmatists will answer Hirschberg's rhetorical question negatively.

What is important here is the fact that Radical Pragmatics regards the problem of the one- and two-sided interpretations of lexical items (see section 1.1.2) as a problem of lexical semantics and pragmatics: it should be resolved at the level of the individual item. Starting from this basic assumption there are three possible solutions: first, the one- and two-sided
interpretations embody an ambiguity; secondly, the lexical items at issue are not ambiguous but are vague, can be used loosely; or finally the one-sided interpretation is the meaning of the item (semantic), while the two-sided interpretation is the result of an implicature (pragmatic). However, there is also another way out of the dilemma: it is possible to abandon the assumption that this is a problem of the semantics and/or pragmatics of individual lexical items. In the chapters to follow I will argue that this problem can also be looked at in a different way: it should be regarded at the level of discourse structure.

2.1.6 Tests for implicatures

Horn (1972, 1989) uses three tests in order to establish the presence of an implicatum. It is possible to eliminate an implicatum by contradicting (canceling) or suspending it, using their non-truth-conditionality as critical property. Or it is possible to make the implicatum, which is normally implicit, explicit by asserting it. Horn identifies several syntactic constructions which serve as schemes for the testing for implicatures.

\[ P_i \text{ is relatively weaker than } P_j \text{ on some scale } P, \text{ so that } P_j(x) \text{ unidirectionally entails } P_i(x) : \]

[19] a. (asserting the implicature)
\[ P_i(x) \text{ but not } P_j(x) \]
\[ \text{just } P_i(x) \text{ (hence, } \neg P_j(x) \text{ for any } P_j > P_i) \]
\[ \text{only} \]

b. (contradicting the implicature)
\[ \text{not just } P_i(x), \text{ but } P_j(x) \text{ (as well)} \]
\[ \text{only} \]
\[ \text{(N.B. the } \text{but of contradiction)} \]
\[ P_i(x) \text{ and what’s more, moreover, in fact}^{26} P_j(x) \]

c. (suspending the implicature)
\[ P_i(x) \text{ if not } P_j(x) \]
\[ P_i(x), \text{ or even } P_j(x) \]
\[ \text{at least } P_i(x) \text{ (and possibly even } P_j(x)) \]

(1972: 39/40)

These sentence frames play a very large role in the development of Horn’s theory. I will discuss them, and some examples, in turn.

A study of the phenomena involved in suspending various inferences forms the starting point for the development of the theory of scalar implicature (Horn 1972: chapter 1). Unfortunately, his descriptions of suspension remain informal: suspending some inference amounts to “rendering it inapplicable” (1972:11), or by suspending “the speaker is explicitly leaving the possibility open that a higher value on the relevant scale obtains, with the

---

26 In this sentence frame, *in fact* expresses something like “in reality”, “what is actually the case”. This contrasts with example (7) in this chapter, where *in fact* is used to express epistemic certainty.
suggestion that his or her knowledge of the actual state of affairs is incomplete.” (1989:235)

The notion is mainly illustrated by a large array of examples, which are mostly formed with the help of the sentence frame A, if not B.27

The test of suspension is used for four different purposes: first, it is used to establish the presence of an implicature, as in (20) and (21):

(20) a. It's possible if not probable that John will leave.
     b. It's probable if not certain that John will leave.

(21) a. at least some if not all
     b. at least possible if not necessary

Secondly, suspension is also used as a test to establish how predicates from the same scale are ordered. This is based on the fact that reversing the order of predicates from a scale leads to pragmatic anomaly28 in the case of suspension. Thus, reversing the scalar predicates in the examples (21a) and (b) gives

(21') a. # at least all if not some
     b. # at least necessary if not possible

Thirdly, the failure of suspension (or cancellation) is used to establish that two predicates are not part of the same scale at all, but that they should be ordered on two corresponding scales, a positive and a negative one.

(22) a. It's warm; in fact, it's hot.
       It's not only warm, it's hot.
       It's warm, if not hot.

27 It is a fact of English that a phrase of the form A, if not B can constitute both a concessive and a suspension. These are nevertheless distinguished from each other by various features. Among these are tense, intonation, vocabulary and the semantics of possible continuations (see Horn 1972:42-45 for discussion).

28 In his (1989), Horn distinguishes three types of anomaly judgements: formal ungrammaticality, marginal status and pragmatic anomaly. These remain entirely intuitive.
b. It's cool; in fact, it's *warm.
   cold
It's not only cool, it's *warm.
   cold
It's cool, if not *warm.
   cold

The scales involved in this example are <boiling, hot, warm> and <freezing, cold, cool>. The fact that It's cool, if not cold is acceptable, whereas It's cool, if not warm is not, is taken as evidence for the fact that cool and cold belong to the same scale, but cool and warm do not. And finally, suspension is used as evidence for the existence of a scale all by itself: the possibility of forming a sentence with the form a, if not b, even if the items concerned do not exhibit any obvious semantic similarities, is regarded as evidence for the existence of a scale. However, this is then pragmatically (locally) instead of semantically (universally) defined. Horn gives several examples, one of which is

(23) The picture of Chiang Kai-Shek that emerges is one that rivals Mussolini, if not Hitler, as the very model of a modern major dictator. (from a review of Sterling Seagrave's The Soong Dynasty)

Horn observes that there are presumably no semantic criteria available for ordering proper names on a scale; "yet we clearly draw the implication from [example (23)] that there does indeed exist a scale on which dictators can be ranked, and furthermore that the Führer clearly outranks (outgrosses?) il Duce on this scale." (1989: 241) The observation that scales can thus be locally, pragmatically established was made earlier by Fauconnier (1975a, 1975b) and Hirschberg (1991).²⁹

A second test which indicates the presence of an implicature is the possibility of cancelling or denying it, i.e. overtly negating it, without a contradiction arising. Horn describes the sentence frames of cancellation as follows: “The environments in [19b] do not just suspend, but CANCEL or BLOCK, the upper-bounding implicatum, with the assertion that a higher value on the scale is in fact known to obtain.” (1989:235) Again, some examples by way of illustration:

(24) a. He ate 3 carrots - in fact, he ate 4.
b. You ate some of the cookies - indeed, you ate all of them.
c. Not only is it possible she’ll win - it’s certain she will.
d. She’s patriotic or quixotic - in fact, she’s both.
e. I’m happy - indeed, I’m ecstatic.
f. It’s warm - in fact, it’s downright hot.
(Horn 1984:20)

(25) It's possible that John left, in fact it's certain/he did leave.

Thirdly, implicatures have as defining characteristic that they can be reinforced, i.e. explicitly asserted “without a sense of anomalous redundancy” (Levinson 1983:120). The

²⁹ Hirschberg (1991) is a commercial edition of her dissertation, which originally appeared in 1985. Thus, Horn (1989) was written later.
motivation that Horn offers for this is that “implicated material is not logically established as true, and it is therefore not redundant to so establish it.” (1972:67) Some examples are

(26) c. It's possible that John left, but not likely.
    It's likely that John left, but not certain.

2.1.7 The epistemic modification

One of the distinguishing characteristics of conversational implicature according to Grice is that they are pragmatic and thus, in Grice's view, non-truth-conditional. A speaker is committed to the truth of his utterances, but not to the truth of his implicatures. This pragmatic nature is mostly signalled by adding an epistemic modification to the representation of an implicatum: as far as the speaker knows or believes. In this section I will present a brief survey of views expressed with respect to this epistemic modification and argue why I will omit it throughout this study.

The epistemic modification is often summarized in the literature, that is, its presence is indicated by a single symbol in front of the sentence that expresses the implicatum. There is however no uniform way of doing this. The abbreviations one finds are, among others, $P$ or $POSS$, for "possibly", $K$, for "the speaker knows", or a combination of these. E.g. Gazdar (1979) and Levinson (1983) express scalar implicatures with $K_p$, "the speaker knows that $p$", and clausal implicatures using $Pp$, "the speaker doesn't know whether $p$" or "it is epistemically possible that $p$" (following Hintikka (1962)). Thus, (27b) is the scalar implicatum of (27a), and (28b) is the (schematical) clausal implicatum of (28a).

(27) a. Some of the students were sick.
    b. $K\neg$ (all of the students were sick).

(28) a. $p$ or $q$
    b. $\{Pp, P\neg p, Pq, P\neg q\}$

The exact phrasing of the epistemic modification has given rise to some discussion; some authors have introduced grades of strength. In reaction to his own definition of scalar implicature of 1972, Horn (1989:234) observes that the epistemic modification should preferably be $POSS \neg$, and "in just those contexts where the addressee assumes that the speaker possesses all the relevant information, $POSS \neg$ can be strengthened to $K\neg$. " (where $POSS \neg$ stands for "for all the speaker knows, it is not the case that", and $K\neg$ stands for the stronger "the speaker knows that it is not the case that"). Levinson in various publications (Atlas and Levinson (1981), Levinson (1983)) claims that scalar implicature differs from the other types of $Q_1$ implicature in the strength of the inference, expressed by a difference in epistemic modification: scalar implicata are modified by the stronger $K\neg$, "the speaker knows that not", while other $Q_1$ implicatures are modified by $\neg K$, "the speaker does not know that". This distinction is not supported by any arguments, however. I find it extremely doubtful whether a distinction like this can consistently be made on the basis of intuition alone.

Hirschberg (1991:79-81) expresses the epistemic uncertainty that implicatures convey in a different way. She proposes that all implicata are disjunctions, formulated in terms of speaker belief rather than speaker knowledge. The implicatum licensed by the affirmation of a scalar value would be "$S$ believes higher $p_j$ are false or $S$ does not know whether higher $p_j$ are
true or false”. She motivates this choice by referring to Grice, who suggested that, as implicata are to some extent undetermined, they should be represented as disjunctions.

As already indicated, I shall omit the epistemic modification in this study. The reason for this lies in the character that I believe Q1 implicature to have. The intuition is very clearly worded by Harnish (1976), who makes a distinction between what he calls direct and indirect implicatures. Indirect implicatures are those that result from the flouting (i.e., deliberately violating) of a maxim by the speaker. These implicatures are dependent on the speaker's intentions: the fact that the implicature is induced depends in part on the fact that the speaker intends to convey it. Indirect implicatures are epistemically modified, they have the form “the speaker means (thinks) that q”. Direct implicatures on the other hand follow from the fact that the speaker is observing the maxims; these implicatures are induced regardless of the speaker's intentions and are not epistemically modified. The reason for introducing this distinction lies in the intuitive characteristics of Q1 implicature, which Harnish identifies as a form of direct conversational implicature: the intuitive inference from the flag is red to "the flag is all red" is different in character from Grice's prototypical examples of implicature. Harnish illustrates the difference by opposing the classical Gricean examples in (29) and (30) to the examples in (31) and (32):

(29) a. S (to H) How is C getting on in his job at the bank?  
   b. H (to S) Quite well, he likes his colleagues and he hasn't been sent to prison yet.  
   (implicated: S's saying (b) implicated that C is the sort of person who might yield to monetary temptation)

(30) x is meeting a woman this evening.  
   (implicated: the woman is not his sister, mother, wife, or close platonic friend)30

(31) x went into a house and found a tortoise inside the front door.  
   (implicated: the house was not x's)

(32) The flag is red.  
   (implicated: the flag is all red)

With respect to the inference induced by (31), Harnish remarks that

In saying I found it in a house, I am not trying to get you to work it out that (I think that) the house is not mine. The implication is carried regardless of this. (361)

In other words, the inferences induced by the examples (31) and (32) are not dependent on any intention on the part of the speaker to convey them. Although I do not agree with Harnish that these inferences are a form of conversational implicature, I take his intuitive characterisation to be an argument to drop the epistemic modification altogether for the exclusion inference.

2.1.8 Summary

The discussion in this section centered around the Gricean argument presented in section 2.1.1. This Gricean argument states all the essential features of (notably) Horn's version of

30 See section 1.2.1 for a discussion of this example.
Radical Pragmatics. Apart from assuming Grice’s maxims, this argument makes specific claims with respect to the semantics of scalar items, the fact that an implicature depends on the availability of some expression alternative, and the nature of the relation between the implicature-inducing item and its expression alternative. Defining the notion of a scale (in any way) can be seen as providing, for a particular set of lexical items/semantic concepts, an answer to the question what the expression alternatives are. Other important aspects of the theory of Radical Pragmatics discussed here are the role of context, which is generally not worked out very well, and the tests used to establish the presence of an implicature.

2.2 Theoretical Problems for Radical Pragmatics

2.2.1 Defining scalar implicature

Very prominent in the discussion on Radical Pragmatics are the definitional problems that this theory suffers from. Most of these were inherited from Grice’s work. As was already pointed out in passing in section 2.1, even the most central notions of Radical Pragmatics remain unsatisfactorily defined. In this subsection I will focus on the definition of Q1 and scalar implicatures. The definition of scales and of what constitutes an expression alternative is the topic of the next subsection.

In 1972, Horn had a “first stab at schematizing the generation of quantity-based implicata” (1972: 90; 1989: 232). This formulation suggests that the definition is intended to cover all implicatures derived by Q1, but this suggestion is contradicted in the first line of the definition itself. It is not clear whether Horn is actually claiming that scalar implicatures exhaust the set of Q1 implicatures:

Given a quantitative scale of n elements, \(<P_n, P_{n-1}, \ldots, P_2, P_1>\), and a speaker uttering a statement \(S\) which contains an element \(P_i\) on this scale, then

(i) the listener can infer NOT \(S(P_i/P_j)\) for all \(P_j>P_i\) (j not equal to n)

(ii) the listener must infer NOT \(S(P_i/P_n)\)

(iii) if \(P_k>P_j>P_i\), then NOT \(S(P_i/P_j)\) implies NOT \(S(P_i/P_k)\)

In (1989), Horn abandoned this attempted definition and chose not to have a second try. The arguments brought in against this definition concern various aspects of it. First of all, the distinction between (i) and (ii), between what a listener can and must infer, is not worked out very clearly. Secondly, as Hirschberg (1991) points out, this definition is not in the spirit of Grice: it regards implicatures in terms of what the listener can or must infer, while Grice introduces implicatures from the perspective of the speaker’s intentions. Finally, and probably

31 Grice's original statement of the theory of conversational implicatures leaves many fundamental issues open to discussion. This concerns e.g. the status of the maxims: Are they rules or not? (the fact that they can be openly violated argues against this). There has been a great deal of discussion over the formulation of the maxims: Aren't Q1 and Relevance the same? Is Quality properly characterized as (only) a maxim? Shouldn't it be an overall principle like the Cooperative Principle? Reformulations of the maxims have been suggested by Harnish (1976), O'Hair (1969), Horn (1984) (see footnote 4, this chapter), Martinich (1980), Levinson (1987). These issues have now become too hackneyed for a fuller discussion. They are, moreover, of only marginal relevance to this study.
most importantly, this definition concerns only quantitative scales, defined by logical and semantic entailment. However, it is clear that various types of nonlogical, pragmatic inference also support scalar implicature.

An entirely different definition is proposed by Gazdar (1979:58/9). He defines a function $f_S$ that has as output the set of potential implicatures dependent on quantitative scales of some utterance (he explicitly excludes nonlogical, pragmatic scales). Later on in the process these potential implicatures pass through a filter (consistency with context) in order to determine which are the actual implicatures of the utterance in question. This is his definition:

$$f_S(\psi) = \{x : x = K^-_{\phi_{a_i}}\}$$

for all $\phi_{a_i}$ such that for some quantitative scale $Q$, $a_i$, $a_{i+1}$ $Q$

(i) $\psi = X \phi_{a_{i+1}} Y$ where $X$ and $Y$ are any expressions, possibly null

(ii) $[\psi]$ subset of $[[\phi_{a_{i+1}}]]$

where $\phi_{a_i}$ and $\phi_{a_{i+1}}$ are simple expression alternatives with respect to $a_i$ and $a_{i+1}$.

This says that $\phi$ scalar-quantity-im-plicates that the speaker knows that it is not the case that $\psi$ if and only if there is some sentence $\psi'$, just like $\psi$ except that it contains a "weaker" scalar expression, and which is entailed by $\phi$ and is either identical to $\phi$ or forms a part of it (e.g., it is one conjunct), subject to the constraint that the scalar expressions are not within the scope of any logical functors in $\psi$ or $\psi'$.

This definition, together with its "prose elucidation", poses some intricate problems. Horn, who cites this definition and its explanation in (1989:233), observes in passing that he has edited the explanation. That is, throughout the explanation (but not the definition!) he has written $\phi$ where Gazdar wrote $\psi$ and $\psi$ where Gazdar has $\phi$. Nevertheless, the explanation is still hard to understand, due to ambiguities in pronominal reference, and to the question which is the correct interpretation of both $\phi$ and $\psi'$, neither of which appear in the definition. A more easily accessible explanation of Gazdar's definition should approximately be

This says that $\psi$ scalar-quantity-im-plicates that the speaker knows that it is not the case that $\phi_{a_i}$ if and only if there is some sentence $\psi'$, just like $\psi$ except that $\psi'$ contains a "stronger" scalar expression. Thus, $\psi$ is entailed by $\psi'$. $\psi$ is either identical to $\phi_{a_{i+1}}$ or $\phi_{a_{i+1}}$ forms a part of $\psi$ (e.g., it is one conjunct), subject to the constraint that the scalar expressions are not within the scope of any logical functors in $\psi$ or $\psi'$.

With respect to the final condition of the definition, "where $\phi_{a_i}$ and $\phi_{a_{i+1}}$ are simple expression alternatives with respect to $a_i$ and $a_{i+1}$", it must be observed that not only should $\phi_{a_i}$ and $\phi_{a_{i+1}}$ be simple expression alternatives with respect to $a_i$ and $a_{i+1}$ (as stated in the definition) but presumably should $\psi$ and $\psi'$ also be simple expression alternatives with respect to $\phi_{a_i}$ and $\phi_{a_{i+1}}$ (as stated in the explanation). After all, the function $f_S$ takes as input the

---

32 $[\psi]$ is the proposition (i.e., the set of possible worlds) denoted by the sentence $\psi$. Condition (ii) amounts to the condition that $\phi_{a_{i+1}}$ entails $\psi$.

33 By definition, "A sentence $\phi$ is simple with respect to an occurrence of a component expression $a_i$ if $\phi$ contains no logical functors having wider scope than $a_i$" (Gazdar 1979:57)
utterance \( \psi \) and gives as a result its potential implicatures, and Gazdar assumes that no implicature is generated when a potentially implicature-inducing expression is in the scope of any logical functor.

Apart from these perplexities, several objections can be raised against this definition. As Gazdar acknowledges himself, in limiting his definition to the logical, quantitative scales he does not capture the entire phenomenon of scalar implicature. Horn (1989) objects, first, that the epistemic condition \( K^- \) is too strong, and secondly (with Hirschberg), that the definition is too restricted in blocking any implicature from an implicature-inducing item which is under the scope of any logical functor. The following counterexamples against Gazdar's definition come from Hirschberg (1991:73):

(33) a. It is possible that Paul ate some of the eggs.
    b. It is possible that Paul ate all of the eggs.
    c. Paul ate some of the eggs or Paul is a liar.
    d. Paul ate all of the eggs or Paul is a liar.
    e. Some people think Paul ate some of the eggs.
    f. Some people think Paul ate all of the eggs.

In (33), the uttering of (a), (c) or (e), respectively, licenses the implicatures \( \neg(b) \), \( \neg(d) \) and \( \neg(f) \). Hirschberg concludes from this that only negation, but not the other logical functors, blocks an implicature.

With respect to the general notion of implicature, Hirschberg makes a different claim from Horn or Gazdar. She does not pretend to define scalar implicature, let alone \( Q_1 \) implicature in general. After characterizing intuitively the inferences she wants to capture in her theory she summarizes the results of this in the form of conventions. She characterizes three subclasses of scalar implicature, namely those arising from affirming, denying or declaring ignorance of some predicate. The implicature may involve a higher, lower or alternate value.

Given an ordering \( O \) that is salient in a context \( C_k \) a value \( v_j \) in \( O \), and an utterance \( u_j \) affirming, denying, or asserting ignorance of \( v_j \).

**Imp1:** If \( u_j \) affirms \( v_j \), then for all \( v_k \) such that \( v_k \) is higher in \( O \) than \( v_j \), \( S \) may license the scalar implicature that \( v_k \) is false or unknown; and, for all \( v_l \) such that \( v_l \) and \( v_j \) are alternate values in \( O \), \( S \) may license the scalar implicature that \( v_l \) is false or unknown.
**Imp2:** If $u_j$ is a denial of $v_j$, then for all $v_l$ such that $v_l$ is lower in $O$ than $v_j$, $S$ may license the scalar implicature that $v_l$ is true or unknown; and, for all $v_l$ such that $v_l$ and $v_j$ are alternate values in $O$, $S$ may license that scalar implicature that $v_l$ is true or unknown.

**Imp3:** If $u_j$ is an assertion of ignorance of $v_j$, then for all $v_l$ on $O$, such that $v_l$ is lower than $v_j$, $S$ may license the scalar implicature that $v_l$ is true or unknown; for all $v_k$ on $O$ such that $v_k$ is higher than $v_j$, $S$ may license the scalar implicature that $v_k$ is false or unknown; and, for all $v_l$ such that $v_l$ and $v_j$ are alternate values in $O$, $S$ may license the scalar implicature that $v_l$ is true.

(1991:65)

This is a very clear characterization of the inferences that Hirschberg regards as scalar implicature, but as she makes clear from the outset, it is not a definition. It is one of the major weaknesses of the theory of Radical Pragmatics that a satisfactory definition of the central phenomenon is lacking.

### 2.2.2 Identifying the expression alternative

It is a general characteristic of Q1 implicatures, and thus of scalar implicatures, that their generation is dependent on the availability of some expression alternative. In the classical version of Radical Pragmatics, exemplified in the Gricean argument in section 2.1.1, this expression alternative is stronger than the utterance actually made (condition (ii) of the Gricean argument). The relative strength of utterances is determined by one-sided entailment. In Hirschberg's elaborated version of implicature theory, the expression alternative is either a stronger, a weaker or an alternate value (see section 2.1.2) on some salient ordering. The problem of identifying the expression alternative, be it a weaker, a stronger or an alternate value, is generally recognized as one of the central problems of the theory of Q1 implicatures. The problem of defining scales satisfactorily is a specific version of this, but does not exhaust it: not all Q1 implicatures are scalar implicatures. After discussing the problem of defining scales, I will briefly discuss two examples of non-scalar Q1 implicatures.

The problem of defining a scale consists of two subproblems: determining, first, what lexical items or concepts are elements of the same scale, and secondly, how these elements are ordered. To start with the second question: it was soon recognized that allowing only one-sided entailment as the ordering relation was much too narrow. Hirschberg widens the perspective substantially and claims that any ordering relation (or metric), linear or hierarchical, which meets some demands, is capable of supporting implicature. The metrics she aims at provide

a way to determine, for any two items ordered in that relation, whether one item is higher or lower than another, or whether the two are alternates with respect to some common higher or lower value.

(Hirschberg 1991:122)

However, even this characterization leaves room for counterexamples. In (34) we have a counterexample both against Horn's and Hirschberg's notion of scale and scalar implicature.

(34) Q. What did you buy for your mother?
A. I bought her flowers.

In (34) we do not want A's utterance to induce the implicature "as far as the speaker knows, he did not buy her tulips". However, Horn's criteria support a quantitative scale <tulip, flower>, as I bought tulips entails I bought flowers but not vice versa. Thus, Horn's criteria do not exclude the implicature mentioned. And even Hirschberg's criteria do not exclude it. Suppose that the salient scale is as mentioned, with as metric kind_of. This metric is accepted by Hirschberg's criteria, and her convention Imp1, cited above, then allows for the implicature "the speaker did not buy his mother tulips". Thus, the notion of a scale is underdetermined both in Horn's and Hirschberg's versions.

The second subproblem of the problem of defining a scale consists in determining which are the members of a scale. Many criteria for scale-membership have been suggested, but all of these have their problems. Generally, one finds two types of criteria, for scales are orderings of semantic concepts and these concepts have lexical realizations. Thus we see both semantic criteria, addressing semantic concepts, and criteria addressing lexical items. To start with the latter category, a necessary criterion within the Gricean theory is that the implicature-inducing item and its expression alternative are equally brief, in order to avoid implicatures by the maxim of Manner. By Manner (first submaxim, "be brief"), elaborate expressions where shorter ones were available lead to an implicature. Uttering the elaborate (35a) in the presence of the shorter expression (b)

(35) a. She produced a series of sounds.
   b. She was singing.

leads to the implicature that what she was doing was not properly singing, that her singing was exceptionally bad. There exist various forms of the brevity condition. Another criterion pertaining to the lexical realizations of the semantic concepts making up a scale is a syntactic one by Levinson (1983:133): the various items on a scale should be "contrastive expressions of the same grammatical category".

Most criteria suggested for scale-membership are semantic in nature, however. Implicitly, they all appeal to the ill-defined notion of a semantic category. Atlas and Levinson (1981:41) suggest that all items on a scale are "about" the same thing, where "about" is a technical term adopted from Putnam (1958). In spite of their explication, it remains a rather mysterious notion. It is an intensional notion; thus to say that sentence A is "about" A does not imply any claim with respect to the reality or existence of the members of set A. But apart from that, nothing positive is said about what it means to say that the sentence A is "about" the set A.

The strategy adopted by Gazdar is characteristic for the problems of definition in this area of Radical Pragmatics. In earlier work, Gazdar had surveyed several criteria which potential members of a scale should meet ("qualitatively similar", "identity of selectional restrictions", "identity of item-induced presuppositions"), but in (1979) he rejects them all. As

---

34 Although in this particular case, an explanation of exactly the same implicature, but by Q1, is perfectly possible as well. Classically, She was singing entails She was producing a series of sounds but not vice versa. Consequently, uttering the second, less informative utterance leads to the implicature that the stronger expression is not valid, as far as the speaker knows.

35 In the philosophy of language, it is important to distinguish the notions intensional and intentional. The former relates to the question of the status of referents, the latter relates to the fact that speech is directed at something in the world. Atlas and Levinson seem to confuse these notions: they write intentional but their explanation is of intensionality.
there are no satisfying definitions available, Gazdar in the end assumes "like Horn, that the scales are, in some sense, "given to us"." (1979:58)

Again, Hirschberg (1991) takes a different strategy from other Radical Pragmatists. She does not suggest any criteria for scale membership, nor does she attempt to define the notion of a scale. Her notion of scale is very liberal, only delimited by the types of metric that support implicature. Thus, in her view,

Posets supporting scalar implicature may be defined over (expressions that denote) entities, actions, attributes, times, places, or concepts, including those concepts ordered in Horn's canonical quantifier, modal, and number scales – or any other items. (1991:125)

Thus we must draw the conclusion that again one of the central notions of Radical Pragmatics, namely that of a scale, is not satisfactorily defined. We said in the beginning of this section that the problem of defining the notion of a quantitative or pragmatic scale is a specific version of the general problem of determining the expression alternatives for some utterance. To finish this section, I will briefly show that this problem exists just as strongly for a subcategory of Q1 implicatures which, according to standard theories, is not a form of scalar implicature. These are the clausal implicatures, a category introduced by Gazdar (1979). That they are indeed a subclass of the Q1 implicatures, without any special characteristics, is very obvious in Levinson's formulation (1983: 136):

Clausal implicatures: If S asserts some complex expression p which
(i) contains an embedded sentence q, and
(ii) p neither entails nor presupposes q and
(iii) there's an alternative expression r of roughly equal brevity which contains q such that r does entail or presuppose q;
then, by asserting p rather than r, S implicates that he doesn’t know whether q is true or false, i.e. he implicates Pq & P q.

The crucial condition is of course (iii): "there is an alternative expression r of roughly equal brevity which contains q such that r does entail or presuppose q". The presence of an expression alternative is essential. All examples concern weaker and stronger items on a ‘scale’ dealing with levels of certainty: e.g. since p, q and if p, (then) q. The first is stronger in the sense that p is being asserted (as true), and consequently q as well. From this difference follows the implicature {Pp, P p, Pq, P q} for if p, (then) q:

(36) a. If John sees me then he will tell Margaret.
    b. I don't know that John will see me.
    c. Since John will see me, he will tell Margaret.

(37) a. My sister is either in the bathroom or in the kitchen.
    b. I don't know that my sister is in the bathroom or that she is in the kitchen.36
    c. I know that my sister is in the bathroom.
    c'. I know that my sister is in the kitchen.

36 A more acceptable English sentence would be I don’t know whether my sister is in the bathroom, Gazdar presumably chose the formulation above to indicate clearly the negation of the expression alternatives included in the phrasing of the implicature.
In (36) and (37), a. is the implicature-inducing utterance, c. (and c') is the expression alternative, and b. expresses (part of) the implicature. The category of clausal implicatures is presumably rather small.

Finally, the problem of determining the expression alternative exists for non-scalar, non-clausal Q1 implicatures as well. This example, due to Grice (1989:37/8), was already discussed in chapter 1:

Anyone who uses a sentence of the form \( X \) is meeting a woman this evening would normally implicate that the person to be met was someone other than X's wife, mother, sister, or perhaps even close platonic friend.

The implicature in this case is due to the fact that the speaker used an indefinite NP where a definite one would have been appropriate. We may assume that the implicit question to be answered here is approximately Whom is X meeting this evening? As Grice puts it, "the speaker has failed to be specific in a way in which he might have been expected to be specific, with the consequence that it is likely to be assumed that he is not in a position to be specific." Clearly, this implicatum again depends crucially on the presence of expression alternatives. In this case there is a relation of one-sided entailment between the expression alternatives, which are denied in the implicatum, and the expression actually used: \( X \) is meeting his wife/mother/sister/(platonic) friend entails, but is not entailed by, \( X \) is meeting a woman.

In short, the supposition present in the Radical Pragmatics theory of Q1 implicatures, which surfaces in the formulation of the Gricean argument, namely that in order to derive such an implicature the available expression alternatives needs to be identified, is a highly problematic one. The reason is that there is no definition of these expression alternatives.

2.2.3 Scalar implicatures and truth conditions

The main reason for Horn to adopt the Gricean notion of implicature is the non-truth-conditionality of implicata. In this section I will discuss some issues concerning implicatures and truth conditions, a topic which is not often dealt with in the Radical Pragmatics literature.

Cohen (1971) presents an argument intended to show that at least one of the inferences Grice analyses as non-truth-conditional implicature actually has truth-conditional effects. This is Cohen’s example:

(38) If the old king died of a heart attack and a republic has been declared, then Tom will be quite content.

If we accept, with Cohen, that the order to the events mentioned in the antecedent is relevant to the truth of the whole sentence, then Gricean implicature cannot account for this intuition. After all, Grice (and Horn with him) takes the semantics of natural language and to be identical to those of the logical operator &. And in logic, reversing the order of the conjuncts does not affect the truth conditions of the formula. Thus, in (34) it should be possible, without any changes to the truth conditions, to reverse the events dying of a heart attack and declaring a republic. However, Cohen argues convincingly that this reversed order may not constitute an adequate description of Tom’s state of mind.
In Cohen’s example we are dealing with two easily distinguishable interpretations of the conjunction: one where \textit{and} is taken as synonymous with \&; and one where \textit{and} is interpreted as “and then”. Grice analyzes these as one interpretation without and one with an implicature, respectively. However, in the case of the generalized Q1 implicatures as understood by Radical Pragmatics, such a distinction cannot be made easily. To take the numerals as example, the interpretation “exactly five” for a sentence like

(39) John has five sheep.

is the combined result of the truth-conditional meaning “at least five” and the generalized Q1 implicature “at most five”. Now this provokes the question what the truth-conditional status of the resulting interpretation “exactly five” is. The implicature is non-truth-conditional, which is supported by examples of suspension and cancellation:

(40) a. John has five sheep, if not six.
    b. John has five sheep, and maybe even more.

In both these examples, the second clause is consistent with the (truth-conditional) meaning of the numeral in the first clause; only the implicature is suspended or cancelled. Presumably, in Radical Pragmatics, the "exactly"-interpretation is non-truth-conditional. But this position can be shown to be false:

(41) Q. How many pupils are there in your class?
    A. 31

In this context, i.e. when the numeral is used as an answer to a \textit{how many}-question (when it is, in the terminology introduced in chapter 1, in comment position), the "exactly"-interpretation is truth conditional. When it turns out the answerer has 33 pupils in his class, he will not have spoken truthfully. Summarizing we can say that the non-truth-conditionality of implicatures, which is their primary characteristic, is not always undisputed.

There is, moreover, a further, rather embarrassing problem. Strictly speaking, implicature theory should maintain that \textit{exactly five} means truth-conditionally “exactly at least five”, since \textit{five} means “at least five”, and \textit{exactly} means “exactly”. How this conclusion is to be avoided is not made clear in the available literature.

A different point with respect to truth conditions is the fact that Grice based his theory on a particular conception of semantics and pragmatics. Everything about the meaning of an utterance that is truth-conditional, i.e. can be described or accounted for by standard first-order\textsuperscript{37} logic, belongs to the domain of semantics. All remaining aspects of the meaning or interpretation of an utterance belong to the domain of pragmatics. In Radical Pragmatics, this view is implicitly assumed. However, this view has been rendered out of date by recent developments. The equations truth-conditional = semantic and non-truth-conditional = pragmatic are no longer generally accepted. E.g., it is recognized that contextually sensitive phenomena may be truth conditional. In various forms of dynamic semantics, such as Discourse Representation Theory (Kamp en Reyle 1993), Dynamic Predicate Logic

\textsuperscript{37} The restriction “first order” is standardly used in the pragmatic literature. It is, however, entirely unmotivated, as it makes no difference at all, for the purpose of the discussions at hand, whether one quantifies over individuals (first order) or over sets (higher orders).
(Groenendijk etc.) and Discourse Semantics (Seuren 1985) we find a conception of semantics that differs radically from the semantics on which Grice based his views. Many aspects of language which were formerly not counted as part of the semantics of language, such as anaphoric reference across sentence borders, are now conceived of as a central semantic phenomenon which any proper semantic theory should be able to account for. In the light of these developments, a critical look at the choices Gricean pragmatics makes with respect to truth conditions and the division of labor between semantics and pragmatics is warranted.

### 2.2.4 Tests for implicature

Sadock (1978) gives some very detailed arguments regarding Grice’s criteria for implicatures, and the problems attaching to them. The arguments presented are equally valid for Radical Pragmatics, as this theory adopted Grice's theoretical foundations without any changes. Sadock notes that only three of Grice’s six criteria are in fact candidates for practical tests. These are calculability, non-detachability and cancellation. However, each of these has its problems; many of the arguments presented here are due to Sadock, but this presentation does not exhaust his collection.38

First of all, calculability is a necessary but not a sufficient condition for being an implicature. The main reason for this lies in the fact that the maxims are much too strong: they allow to derive almost anything from anything. Moreover, there is a considerable amount of redundancy or overlap in the maxims (as Grice was well aware), which makes the choice of a particular maxim for deriving a particular implicatum very hard (see e.g. the example in footnote 24 of this chapter). Thus, calculability cannot be said to be unequivocal. The most notable case is that of Relevance and Quantity: isn't "not saying more or less than is required" simply the same as being relevant?

Secondly, non-detachability is neither a necessary nor a sufficient condition. That it is not a necessary condition was readily acknowledged by Grice himself, because of the clash with the maxim of Manner. In order to test for non-detachability it is necessary to find two ways of expressing the same content, to be able to check whether they carry the same implicata. However, if these two expressions are not equivalent in many respects, e.g. if one is much longer than the other, then this will induce an implicature by the submaxim "be brief" of Manner. Consequently, it is no longer possible to compare these two expressions with respect to e.g. only their Quality-implicata. Non-detachability is not a sufficient condition either. The main reason for this lies in the fact that non-detachability does not distinguish between implicature and entailment (see also Grice 1989:86-116). This is easily understood: necessarily, two utterances expressing the same content carry the same entailments, just as they would induce the same implicata. Apart from these theoretical counterarguments against non-detachability, there are also two practical arguments. It is often very hard to decide whether two expressions have the same meaning, which is a necessary precondition for successfully applying the test. Moreover, there are many cases where the application of this criterion is severely limited, due to the fact that there are no synonymous expressions that avoid the clash with the maxim of Manner (see also section 2.1.1).

---

38 In this discussion, I have omitted Sadock's arguments pertaining to the difference between conversational and conventional implicatures. Sadock regards presuppositions as a form of conventional implicature, and as presuppositions are not the topic of this dissertation, I will not discuss them here. Some brief remarks on the difference between conversational implicatures and presuppositions are made in section 2.1.1 and 2.2.4.1.
However, the most serious problem that Sadock points out regards cancellability. This is generally considered to be the best of the tests, and as we saw in section 2.1.6 it plays a very central role in justifying the theory of Radical Pragmatics. Sadock's argument concerns the very core of Radical Pragmatics, which was designed (among other things) to get rid of unwanted (lexical) ambiguities. But Sadock claims that "the test [of cancellability] does not distinguish cases of ambiguity from cases of univocality plus possible conversational implicature." (293) His argument is that when we have a grammatically ambiguous sentence, such as (42a), it is always possible to contradict (i.e., cancel) one of the readings, thereby leaving only one available (as in (38b)):

(42) a. Everyone speaks one language.
   b. Everyone speaks one language although no one language is spoken by everyone.

This is a serious problem for Radical Pragmatics, for “the sad fact is that in the very cases where argument is likely to arise as to whether something conveyed by an utterance is conversationally implicated, the competing claim would be that the utterance is ambiguous.” Thus, on the basis of cancellability it is not possible to decide whether e.g. an utterance containing a numeral is ambiguous or whether it has a unitary meaning and induces a generalized implicature.

The final test that Sadock discusses is reinforceability. This test is based on the fact that conversational implicatures are not part of the conventional meaning of an utterance. Therefore, "it should be possible to make them explicit without being guilty of redundancy." E.g., it is perfectly acceptable to say

(43) Mary ate some, but not all, of the cheese.

Reinforceability is subject to the same objections as was cancellability: it does not distinguish between an ambiguity analysis and an analysis in terms of conversational implicature. Thus, neither cancellability nor reinforceability can be used as an argument for an implicature analysis, against an ambiguity analysis. In (34c), the ambiguity is resolved by reinforcement of one of the two possible readings:

(42) c. Everyone speaks one language and it is the same language.

2.2.4.1 Implicatures and presuppositions
Incidentally, if we accept reinforcement as a criterion for conversational implicature, then this provides us with a criterion for distinguishing implicatures and presuppositions from each other. The crucial point is that, as Sadock observes, "the reinforcement test [for implicatures] is sensitive to the order of the reinforcing expression and the implicature-bearing expression." (294) That is, reinforcement of an implicatum only produces an acceptable utterance when the reinforcing expression follows the utterance inducing the implicature (as in (43) above), and not when it precedes it, as in (43'):

(43') ??But not all, Mary ate some of the cheese.
On the other hand, presuppositions can be made explicit as well, but then an acceptable piece of discourse only ensues when the presupposition is uttered before its bearer. Observations leading to this conclusion have been made by several authors, among others by Seuren (1972: 360), and by Karttunen and Peters on several occasions (1975, 1979:14 "ideally, every [presupposition] ought to belong to the common set of presumptions that the utterance of the sentence is intended to increment.").

(44) a. The present queen of the Netherlands usually wears a hat.
   b. There is a queen of the Netherlands/ the Netherlands have a queen.

(45) a. The Netherlands have a queen and she usually wears a hat.
   b. ?The present queen of the Netherlands usually wears a hat, and the Netherlands have one.

(44b) represents the (existential) presupposition of (44a). In (45a), the presupposition is made explicit and an acceptable piece of discourse is formed: the presupposition precedes its bearer-sentence. (45b), with reversed order, is not acceptable, at least on an interpretation in which one and present queen of the Netherlands are coreferential (if they are not coreferential we do not have a presupposition here). This difference in distribution between the explicit realizations of presuppositions and implicatures is a reflection of their difference in nature.

2.2.5 Summary

The theoretical problems for Radical Pragmatics are predominantly problems of definition. Generally accepted definitions of central notions such as (scalar) implicature, scale and expression alternative are lacking. As a result, there is no agreement on what is the best method to select the relevant expression alternative at any point in a discourse. Hirschberg avoids these problems by presenting only summaries of detailed empirical surveys. Furthermore, attention was drawn to the role of truth conditions, and to the division of labor between semantics and pragmatics implicit in Radical Pragmatics. Recent developments in semantics provide an argument to question this division of labor, in which everything which is truth-conditional belongs to the field of semantics, and everything which is non-truth-conditional belongs to the field of pragmatics. Finally, the various tests for implicature are not unproblematic; theoretical as well as practical counterarguments were presented. Following Sadock's discussion of the test of reinforcement, we were able to formulate a new criterion for distinguishing implicatures and presuppositions. Although both types of inference are normally implicit and can be formulated explicitly, there are differences in the acceptability of these explicitations. A concatenation of an (explicit) presupposition and the presupposition-bearing utterance is only a well-formed piece of discourse if the presupposition precedes its bearer-sentence. But a concatenation of an (explicit) implicature and the implicature-inducing utterance is only well-formed if the implicatum follows its inducing utterance.
2.3 Observational Issues

2.3.1 Entailment

Much text has already been devoted to the role of entailment within the theory of Radical Pragmatics, e.g. as to whether it is a necessary and/or sufficient criterion for forming a scale, and whether or not it is the only relation capable of forming a scale. In this section I will not discuss these theoretical issues but make some observations which contradict (notably Horn's early version of) Radical Pragmatics. Intuitive judgments on various aspects of language form the foundation of linguistics, and determining a relation of entailment is among these intuitive judgments.

First of all, as Hirschberg observes, the relation of one-sided entailment that Horn claims to exist between all and some is only valid in Aristotelian logic, where empty sets are not allowed. In present-day logic there is no such assumption, and thus no relation of one-sided entailment between all and some. This removes the foundation for the scale <all, most, many, some>. But apart from this logical or semantical argument against positing this scale, there is also an argument pertaining to the discourse function of the predicates all and some. From the perspective of their function in discourse, all and some are not completely the same. Some has two functions: introducing new individuals (discourse referents) into the discourse, as in (46):

(46) I saw some boys entering the room,

or referring to a (strict) subset of some (previously introduced) set. This anaphoric use is exemplified in (47):

(47) a. There were 15 boys at the party. Some of them left early.
   b. Some of the boys who were at the party left early.

Only in this latter use does some convey "not all". All, on the other hand, functions only anaphorically, referring back to (the members of) a previously introduced set (as in (48a)).

(48) a. There were 15 boys at the party. They all left early.
   b. I saw all (of the) boys entering the room.

If there is no such set, as in (48b), it is accommodated post hoc; or in other words: all carries an existential presupposition.

Incidentally, this presuppositional analysis of all (which is observed in numerous places, among others in Strawson (1952:174)) supports the existence of an entailment relation from all to some. This is further supported by the unacceptable flavor of an example like

(49) !Maybe John has no children, and then all his children are asleep.39

What we find here, with respect to all, is a contradiction between a conclusion based on logical considerations (namely, the entailment relation is not logically valid) on the one hand, and a conclusion based on considerations pertaining to the discourse function of lexical items on the other hand. This dissertation is certainly not the place to resolve this contradiction; but it does seem justified to conclude that Radical Pragmatics is acting too hastily. Radical

39 The test employed here is explained in detail further down in this section.
Pragmatics assumes the semantics of *some* and *all* to be synonymous to that of the logical operators; in that case, the objection raised by Hirschberg is a very serious one. But if one, in trying to avoid this objection, adopts a different view and allows for considerations pertaining to the discourse function of these items to enter the discussion, then, due to these differences in function between *some* and *all*, one cannot just treat them on a par, as is done in Radical Pragmatics. Only in their anaphoric use does there appear to be a relation between them.

There are more scales which pose problems in the light of the entailment-condition. The fact that a scale such as <hot, warm, tepid, cool, cold> cannot be defined in terms of one-sided entailment (*This soup is hot* does not entail *This soup is cold*) induced Horn (1989:243) to divide this and similar scales into a positive and a negative pole: <(boiling), hot, warm> and <(freezing), cold, cool>. Intermediate values such as tepid and lukewarm belong to both scales and thus connect them. Graphically, these so-called extended scales can be represented as in (50):

![Graphical representation of scale](image)

A different type of problem arises in an example like (51) (Hirschberg (1991:50):

(51) A. Did you get Paul Newman's autograph?
B. I got Joanne Woodward's.

In the former case it was necessary to divide the scale of temperatures into two halves in order to save the entailment-condition, but in this example there is no entailment-relation at all. Still, proper names like this are capable of inducing implicatures, as both Horn and Hirschberg observe. In other words, the entailment-condition is not a necessary condition for forming a scale.

Although this conclusion severely limits the importance of one-sided entailment as the ordering relation in a scale, there are two more observations to be made here. First of all, an ordering with one-sided entailment as the ordering metric is supposedly an ordering by informativity. The entailing expression (e.g., *the soup is boiling*) is considered stronger or more informative than the entailed expression (e.g., *the soup is hot*). In general this accords quite nicely with intuitions, but there are situations where one-sided entailment imposes an ordering according to informativity where, intuitively speaking, there is no such ordering. This is most obvious in cases which involve numerals. According to standard Radical Pragmatics, it is always more informative to utter *John has four children* than it would be to utter *John has three children*. But as I will argue in more detail in chapter 3, in a context where the number of John's children (or more specifically, the cardinality of the set of John's children) is at issue, giving a higher number by itself does not count as being more informative.
The second observation to be made here considers the list Horn gives in order to illustrate his intuitive notion of a scale. The list was quoted in section 2.1.2, repeated here:

<always, usually, often, sometimes>  
<. . . , 6, 5, 4, 3, 2, 1>  
<necessary, (logically) possible>  
<obligatory, permitted>  
<freezing, cold, cool, (lukewarm)>  
<hideous, ugly, unattractive, plain>  
<loathe, hate, dislike>  
<{terrible/awful}, bad, mediocre>

With respect to examples like these, Hirschberg asked the rhetorical question “It is easy to say that beautiful does NOT entail hideous or ugly – but does it entail plain – or even pretty for that matter?” (1991:116). I would like to pursue this line here and show that, intuitively, a relation of entailment does not always exist among sentences containing the proper lexical items mentioned above. In logic, entailment is defined as follows: A entails B (A |= B) just in case it is analytically impossible for A to be true and B to be false at the same time. This definition has an observational counterpart: “when A entails B, a speaker cannot at the same time assert A and explicitly leave open the possibility that not-B” (Seuren 1991:290 (English version by the author)). Thus, when a speaker combines the possibility of not-B with A, while A entails B, this should lead to an intuition of contradiction (signalled by an exclamation mark in front of the sentence). A clear case is (45), employing the example <tulip, flower> which already appeared in section 2.2.2:

(53) a. He bought his mother tulips |= He bought his mother flowers.  
b. !Maybe he didn't buy his mother flowers, but he did buy her tulips.

On the other hand, there is no entailment relation between two sentences when the combination of "possibly not-B" with A does not lead to a sense of contradiction, as in

(54) a. He didn't buy his mother tulips |≠ He bought his mother flowers.  
b. Maybe he didn't buy his mother flowers, but he certainly did not buy her tulips.

Now this test for entailment can be applied to Horn's examples of scalar items; in the examples to follow, in (a) and (b) I will list the sentences at issue, such that (a) supposedly entails (b). The sentence in (c) is the test-sentence. The result of this test turns out to be that the whole class of scalar items as assumed by Horn divides up into two subclasses: the scalar adjectives on the one hand should be distinguished from the numerals and the quantifiers on the other hand. Let's start with the latter category:

(55) a. John has at least four children.  
b. John has at least three children.  
c. !Maybe John doesn't have at least three children, but he definitely has at least four.
In this example, the presence of an entailment-relation between the (a)- and (b)-sentence is straightforwardly affirmed by the test-sentence. However, as I will show in chapter 3, this is compatible not only with a Radical Pragmatics-theory about the interpretations of numerals, but also with a different theory. The presence of this entailment-relation does not force us to assume an explanation in terms of implicatures. Examples employing the quantifier-scales also support Horn's claim:

(56) a. It is obligatory to do your assignments.
    b. You are permitted to do your assignments.
    c. !Maybe you are not permitted to do your assignments, but it is obligatory to do them.

(57) a. You must clean up your room.
    b. You should clean up your room.
    c. !Maybe you shouldn't clean up your room, but you must do it.

The second category is that of the adjectives. In this category, by the test employed here, it is often hard to detect an entailment-relation.

(58) a. I adore him.
    b. I love him.
    c. Maybe I don't love him, but I do adore him.

(59) a. I adore him.
    b. I like him.
    c. Maybe I don't like him, but I do adore him.

(60) a. He is hideous.
    b. He is plain.
    c. Maybe he isn't plain, but he certainly is hideous.

(61) a. She is beautiful.
    b. She is plain.
    c. Maybe she isn't plain, but she certainly is beautiful.

(62) a. She is beautiful.
    b. She is attractive.
    c. Maybe she's not attractive, but she sure is beautiful.

(63) a. She is beautiful.
    b. She is pretty.
    c. Maybe she's not pretty, but she definitely is beautiful.

(64) a. His work is terrible.
    b. His work is mediocre.
    c. Maybe his work isn't mediocre, but it's simply terrible.
On the test applied here, there is no relation of entailment between the (a)- and the (b)-sentences in any of these examples. The theory of scalar implicatures involves a strong claim about the semantics of the scalar adjectives involved. Various items on a scale are assumed to designate different grades of intensity of the same property. However, from the examples given above it can be concluded that other interpretations of these adjectives are possible. E.g., from (54) it can be concluded that beautiful and attractive designate two radically different properties, not different grades of intensity of the same property. It is perfectly possible to find a person beautiful, though unattractive. As a further example, consider an imposing elderly lady: one may find her beautiful, but not pretty. And in (63) it is possible to find a coherent interpretation along the following lines: pretty applies only to a person's physical features, but beautiful can also be interpreted in a moral sense, i.e. "being a good person". The same goes for the examples in (58) and (59): liking or loving on the one hand and adoring on the other express separate emotions, not grades of intensity of the same emotion. Similar processes work in the other examples employing Horn's scalar adjectives. E.g. in (64), the qualifications mediocre and terrible can be understood to refer to two different aspects of the work at issue, e.g. it's scientific qualities and it's stylistic properties.

The interpretations suggested here for the scalar adjectives may not be the most likely ones, or the easiest, but they are possible and coherent, thus denying the presence of an entailment-relation between the (a)- and (b)-sentences. In order to understand properly what happens in the examples above, it is necessary to take a closer look at the sentence frame of the test for entailment. Its form Maybe not ..., but ... suggests (to the hearer) that there is some opposition between the first and the second clause. In other words, a contradiction is built into it, and in the interpretation of sentences formed by it we can detect a strategy aimed at avoiding the impending contradiction. This is best illustrated by way of some examples:

(65) a. He may not be my best student, but he definitely is the smartest.
    b. He may not be my smartest student, but he definitely is the best.

The crucial process is this: the NP the best student may, if not accompanied by any further indication, denote the person from a class who gets the highest grades (achieves the best results). It is common to assume that this person therefore is the smartest pupil as well. Thus, the two notions "being the best student" and "being the smartest student", though not synonymous, are closely related. But in the examples given here (employing the test for entailment) the attention is directed towards their differences. Thus, "being the best student" will get interpreted as "(only) getting the highest grades" and "being the smartest student" as "having the highest/largest intelligence". The result of thus letting the interpretations of the two contrasted notions diverge is that the impending (logical, semantical) contradiction is avoided. In general, what we see if we employ this sentence frame is that the interpretation of semantically similar notions is such that the differences are stressed, in order to avoid a contradiction. A further example may illustrate this tendency:

(66) Ard Schenk may not be a good skater, but he is the world champion.

Under normal circumstances, "being a good (i.e., fast) skater" and "being the world champion" are not necessarily in opposition. However, the form of the sentence (maybe not ..., but ...) suggests that there is some opposition, and the interpretation of these notions is accomodated to reveal it, thereby avoiding the logical contradiction. In this example, a good skater will not be interpreted as "a fast skater" but "a skater having a good technique".
These considerations with respect to the test for entailment, and the processes of interpretation it invokes, provide further support for the adequacy of the test. It will only reveal contradictions which are somehow “central” to the notions employed, not those which are due to one specific interpretation of a term which allows for a multitude of interpretations.

From these considerations two conclusions can be drawn with respect to the class of scalar items that Horn assumes. First, this class is not homogeneous: with respect to this test, numerals and quantifiers behave similarly and thus form one category, but the scalar adjectives are a separate category. Secondly, the relation of one-sided entailment that Horn claims to exist among the concepts on one scale is hard to establish, specifically for the scalar adjectives. When a contradiction looms large, it is possible to interpret these adjectives in such a way that this contradiction is avoided, and thus that the entailment relation cannot be established. The test employed here does not give any clues with respect to the cause of these diverging interpretations. Its results are compatible both with a theory which assumes a narrow semantics combined with pragmatic inferences, and with a theory which assumes a much wider lexical semantics.

2.3.2 Tests for implicature

The fact that the tests for implicature are primarily intuitive in nature, and that they are in many ways not adequate, was discussed above. In this section I want to draw attention to one aspect of the examples of cancellation, namely that they are often only marginally or not at all acceptable. Although both Horn and Gazdar signal this, they do not seem to consider it a grave problem. Horn (1972:95/6) writes
(67) It's possible that John left, and (in fact) he did.40

If the implicit contravention of the implicature is made explicit, as in [62], no logical inconsistency ensues, although we may wonder why the speaker bothered to assert the first conjunct rather than merely entail it by uttering *John left.*

Similarly, with respect to example (68), Gazdar writes

(68) a. My sister is either in the kitchen or in the bathroom.
   c. My sister is either in the kitchen or in the bathroom, and I know which.
   d. My sister is in the kitchen.
   d'. My sister is in the bathroom.

In the c-sentences, we have observed that implicatures can be intrasententially cancelled. One might wonder why this should be permissible, since if one wanted to make the stronger claim in the first place, then a d-sentence would surely do the job better than the longer and potentially confusing c-sentence. However there are cases where a c-sentence allows one to say things not readily sayable with a d-sentence. Thus, [68] could be used by one participant to another during a game of hide and seek. (1979:52)

Neither author suggests an explanation for what causes the marginal status of these examples of cancellation. However, in chapter 1 I proposed to view context in the form of a question, and this method allows us to identify the oddity of this type of example. If every utterance in discourse is the answer to some question, one may try to reconstruct this (or these) question(s). But for example (67) it is not possible to construct a question to which it in its entirety forms an acceptable answer. This is a more exact way of expressing Horn's intuition that "we may wonder why the speaker bothered" to express himself so elaborately. With respect to example (68) Gazdar suggests a possible context in which (68c) could be acceptable: in the context of a game of hide and seek. The implicit question could be *where is your sister hiding?*

In other works on Radical Pragmatics we find many more of these examples, and regarding them in the context of question resolves some of the mystery. Some further illustrations come from Levinson (1983:115/120/138):

(69) a. John has three cows, in fact ten.
   b. Some of the boys went to the party, in fact all.
   c. This soup is warm, in fact hot.

The examples (67-69) together cover the main categories of scalar predicates (quantifiers, logical operators, modal operators, numerals, scalar adjectives), which indicates that the problem addressed here is a problem of the process of cancellation, not one that is dependent on the particular scalar predicate used.

For example (69a) there are two possibilities of supplying a context, one which leads to an acceptable piece of discourse, one which does not (as introduced in chapter 1, the comment of an answer is underlined):

40 The question mark signals marginal status (see footnote 19, this chapter).
Q. How many cows does John have?
A. * John has three cows, in fact ten.

Q. Who has three cows?
A1. John has three cows, in fact ten.
A2. John does, in fact he has ten.

When we encounter *John has three cows, in fact ten in isolation, as in (69a), it is only marginally acceptable; a reason could be that a context as in (70), i.e. a *how many-question, is immediately supplied. This context would be the default context. The context sketched in (71) would lead to a much more acceptable piece of discourse, but it would not be the default context. The relevant difference between (70) and (71) is that in (70), the numeral *three is in comment position, while in (71) it is part of the topic-expression. This difference in structural position of implicature-inducing items will (in the next chapters) turn out to be essential in determining their interpretation.

Example (69b) is, with respect to the question that would form an appropriate context for it, comparable to (67). When one tries to look at this sentence in an unprejudiced way, it turns out that it is awkward. What is the speaker trying to convey? If it is the fact that all of the boys went to the party, then why does he express it so elaborately? As in (67), there is no single question to which (69b) would be a satisfactory answer. The same conclusion can be drawn with respect to (69c). However, something more can be said about both (69b) and (c). Levinson gives these examples as a further argument for the treatment of scalar predicates with generalized implicature: that this treatment is correct becomes clear from the fact that (69c) is not contradictory, as this book is short, in fact long is. But trying to supply a context for (69c) the following small discourse as a result:

(72) <What is the temperature of this soup?>
   It is warm.
   <Are you sure?>
   in fact it is hot.

The only way to construct an acceptable piece of discourse out of the sentence in (65c) is to regard it as two separate contributions to a discourse, notably two different answers to the same question, the second answer being induced by a request for confirmation. In chapter 3 I will argue that this should be regarded as a form of (truth conditional) correction (repair) or further specification. Along the same lines an acceptable piece of discourse can be constructed for (69a) and (b). This method of constructing an acceptable piece of discourse is not so easily available for example (67) It is in general hard to find a question which moves possible into comment position.

Summarizing we can state that the marginal status of many examples of cancellation can be attributed to the fact that they cannot function as an appropriate answer to a question, and thus not as a contribution to a discourse. The anomaly signalled here relates to discourse structure. Some of the examples discussed allow a coherent analysis, but only when the two parts are regarded as separate, truth-conditionally distinct, contributions to a discourse, i.e. as answers to two separate questions.
2.3.3 Inadequacy of predictions

In this section I will present some arguments\textsuperscript{41} that aim to show that Radical Pragmatics cannot work in principle. That is, even if all problems indicated above were solved, if the theory were extended with a theory of context and of relevance, it still would not work. The problem, often observed before, is that Radical Pragmatics can, with equal ease, produce competing explanations for one interpretation (see also note 25, this chapter), or competing interpretations for one utterance. These problems are most easily illustrated by examples involving the numerals.

As we saw earlier in this chapter, Radical Pragmatics invokes implicatures to account for the "exactly"-interpretation of numerals. But an entirely different interpretation of those numerals, namely "at least", can be reached through the same line of reasoning. Suppose that the topic of conversation is the number of children John has. In this context, a speaker utters

(73) John has three children.

Then it is possible to argue that the speaker has conveyed that, as far as he believes or knows, John has at least three children. For we can posit a scale <exactly three, three>, which satisfies the entailment-condition. Exactly three counts as stronger and more informative in the standard version of Radical Pragmatics. From the fact that the speaker didn't utter exactly three, we infer that he is not in a position to utter it, thus that he does not believe that exactly three is true. The speaker knows no more than that three ("at least three") is true. If Radical Pragmatics wants to block this inference, it will have to explain why in this situation the scale < .. 6, 5, 4, 3, .. > is preferred, by speaker and hearer, over <exactly three, three>. This involves, next to positing an ordering over semantic concepts in scales, positing an ordering over scales as well. Considering that the former matter is far from solved, the latter will not be easily solved either.

An additional problem with this line of reasoning is that it excludes the answer at least three to the question how many children John has. It is obvious that this answer is perfectly in order and can thus not be dismissed. In Radical Pragmatics, three and at least three are synonymous. Consequently, Manner forces us to use the least elaborate expression; if we used the longer one, then this would induce implicatures of its own. However, there is nothing odd or unusual about answering at least X to a how many-question; nor does it induce any particular implicature (apart from "the speaker doesn’t know the exact number of children John has"). This example clearly shows the problematic nature of the assumption that three and at least three are synonymous.

Gricean reasoning in this area predicts that, to the question how many children John has, three and exactly three are the correct, appropriate answers, but at least three is not. However, the facts are the other way around: at least three and three are appropriate, exactly three is not. In chapter 3 I will propose an explanation for this observation: in the context at issue here, where the number of John's children is topic, three and exactly three are synonymous.

\textsuperscript{41} J. Groenendijk (p.c.) drew my attention to these arguments.
3 Numerals in Context

3.1 Introduction

In this and the following chapter, I will be arguing the following claim: all examples of exclusion inference (or Q1 implicature) in the literature, of which the examples involving scalar terms are a subclass, involve exhaustive interpretation. This claim will be argued for in two steps: in this chapter, the interpretation of numerals is in focus, in chapter 4 other categories of scalar predicates will be reviewed. This distinction is made for reasons of exposition; the cardinal numerals provide a very clear illustration of the claims I will be making. Thus, from the distinction between numerals on the one hand and other scalar predicates on the other hand, no conclusions with respect to a difference in nature can be drawn. On the contrary, the conclusion from this and the following chapter will rather be that, with respect to the process of exhaustive interpretation, there is no separate category of scalar predicates at all.

In any semantic study it is important to distinguish the lexical item under consideration and its denotation. The lexical items which are the object of study in this chapter will be referred to as numerals, and their denotation as cardinal numbers. There are many sides to the use of numerals in natural language, but I will discuss only one of these: the situation in which numerals are used to indicate a cardinal number. The English word number is ambiguous: it can either refer to a written symbol, or to the cardinality of a set. It is numbers in the latter meaning, i.e. as indicators of the cardinality of a set, which are the topic of this chapter.

Numerals are often treated as the prime example of scalar implicature; e.g. Kempson calls them "one of the core cases of scalar implicature" (1986:86). Implicature theory rests its theory of cardinal numbers on the assumption that cardinal numbers have two readings, the "at least" and the "exactly" reading. I will briefly recapitulate some examples of this and their purported explanation. Next I will discuss a suggested third reading for numerals, namely the "at most" reading.

Some examples may serve to illustrate the "at least" and the "exactly" readings again:

(1) the "exactly" reading
   a. John has four cows.
   b. John bought three books.
   c. Ten children from Jordania are representing their country at this conference.

(2) the "at least" reading
   Q Could you lend me £10?
   A Sure I could lend you £10.

In (2), it is in the answer to the question that the numeral can receive what is often referred to as an "at least" interpretation. This informal characterisation of the interpretation of the numeral in (2A) is in fact rather sloppy. A more adequate characterisation is that, by the way the numeral is used in (2A), the possibility of a higher number (cardinality) obtaining is not excluded. On the other hand, on an "exactly" interpretation, the possibility of a higher number
obtaining is in fact excluded. However, for ease of exposition I will adhere to the common
indicators "at least" and "at most".

Assuming ambiguity for all numerals is an unattractive strategy from the perspective of
scientific parsimony (Grice calls this Modified Occam's Razor: "Senses are not to be
multiplied beyond necessity" (1989:47)). Implicature theory, as we saw in detail in chapter 2,
takes "at least \( n \)" as the truth-conditional meaning for the numerals, and adds a generalized Q1
implicature "at most \( n \)". These together result in the desired interpretation for the examples in
(1). The examples under (2) would be the result of a contextual cancellation of the
implicature.

Apart from the "at least" and "exactly" readings, Carston (1985:10) suggests that
numerals (or other implicature inducing items, as in (3c)) have as a third option an “at most”
reading:

(3) a. She can have 2000 calories a day without putting on weight.
   b. The council houses are big enough for families with three kids.
   c. You can have half the cake.

These examples have in common that the quantities given (2000 calories, three kids, half the
cake) function as an upper limit. Carston presents these cases as one more problem for the
Radical Pragmatics theory of cardinal numbers. However, regarding the "at most" reading as
a third possible interpretation for numerals is not the only possible line with respect to these
examples. The possibility that the "at most" reading is forced by the semantics of other lexical
items in these sentences is not considered. Both can and enough are modal predicates that
involve some implication of an upper bound, which may lead to the “at most” reading for the
numerals. For (3a) and (b), one could supply contexts like the following:

(3)' a. Q How many calories a day can she have without putting on weight?
   A She can have 2000 calories a day without putting on weight.
   b. Q How many people can live in a council house?
      How big are the council houses?
   A The council houses are big enough for families with three kids.

The numerals are used to indicate a cardinality, and the fact that this cardinality represents an
upper bound does not follow from the numeral by itself, but results from the main sentence
predicate in these cases.

Remarkably, a similar point can be made with respect to many of the examples involving
an "at least" reading. These are two examples from Carston (1985:10), in which the semantics
of both have and need should be taken into account when studying the interpretation of the
numeral:

(4) a. In Britain you have to be 17 to drive a motorbike and 18 to drive a car.
   b. Mary needs three A's to get into Oxford.

Another example comes from Gazdar (1979:137/8):

Now [5] implicates\(^{42}\) that the speaker knows that Mrs. Smith does not have more than three children (any number greater than 3 will precede 3 on the quantitative scale in which 3 appears), but there are plausible situations in which this implicature will get contextually cancelled. Imagine that we are discussing the eligibility of each of a group of women for welfare benefits and that we know of each how many children she has (Mrs. Smith, for example, has five). If one of the welfare benefit eligibility criteria is having three children, then one could utter [5] in such a discussion without having been heard as implicating that Mrs. Smith did not have more than three children. [my italics, RS.]

Gazdar claims in fact that (5) can be interpreted as "Mrs. Smith has at least three children", by a process of implicit (contextual) implicature cancellation. But another way of viewing this example would be to say that utterance (5), in this particular context, does not answer the question "How many children does Mrs. Smith have?", but instead the question "Does Mrs. Smith meet the requirements for receiving welfare?" In this view, the semantics of criteria or requirements is responsible for the fact that, on the basis of (5), the possibility that Mrs. Smith has more than three children cannot be excluded. In general, modal predicates like need, must, necessary, requirements and criteria indicate that the amount specified constitutes a minimum requirement. Thus a higher number is irrelevant in these situations. (This holds not only for numbers, but for all possible properties which function as a necessary condition, a requirement etc.) Numerals which are in the scope of a modal predicate can thus be seen to receive an interpretation which cannot be accounted for without taking the semantics of the modal predicate into account. These cases fall outside the scope of the study conducted in this chapter.

According to the view that will be developed in this chapter, it is not the word three that allows for different readings, as Gazdar, with the other Radical Pragmatists, claims, but it is the sentence which is ambiguous between at least two different readings. In different contexts, sentence (5) can be used to make different utterances, and these utterances have different truth conditions. Moreover, one of the fundamental assumptions on which the implicature analysis is based, is faulty. This concerns the assumption that the semantics of numerals can always be adequately accounted for in terms of the existential quantifier. This assumption is common in logic and is responsible for the claim that the truth-conditional meaning of any numeral n is “at least n”. Now this claim is arguably wrong; when a numeral is in comment-position, its truth-conditional interpretation is “exactly n”, or more precisely: the possibility of any other number, higher or lower, obtaining is truth-conditionally excluded. This is the case when a numeral is used to specify a cardinality, to name the number of elements in some specific set, and this use of a numeral is better accounted for as a form of predication. In section 3.4 I will discuss two formal semantic views on numerals and in 3.5 I will discuss the various uses of (cardinal) numerals in natural language and address the question how these can be accounted for by a unitary semantics and pragmatics.

Integration of context is essential to my survey of the linguistic facts in this and the next chapter, and as discussed in chapter 1, I will represent the immediate linguistic context of an

\(^{42}\) In his dissertation, Gazdar builds a system for generating several semantic and pragmatic inferences (presuppositions, implicatures) for a sentence. This system works in two steps: first the potential inferences of a sentence are generated, next these are added to the context in a certain order, during which process contradicting inferences are filtered out. The potential inferences are designated as implicatures, pre-suppositions, the actual (remaining) inferences lose their hyphen.
utterance in the form of a question. Moreover, the adoption of DTT, which is a theory of discourse structure in terms of questions and answers, allows us to distinguish various nuances within the class of exclusion inferences. As I will show in detail in sections 3.5 and 4.2, there is a correlation between the position an implicature-inducing item occupies in the information structure of a discourse (i.e., whether it is part of the topic-expression or the comment-expression) and the status of the exclusion inference it induces. This correlation has the following form: exclusion inferences induced by linguistic items in comment-position are truth-conditional, contrary to what most theories on implicatures claim. When a linguistic item triggering an exclusion inference is part of the topic-expression, the exclusion inference is no more than an invited inference, easily overridden by contradicting information. Or, to formulate this claim independently of implicature theory: all information given by lexical items in comment position induces an exclusion inference. This is not a type of inference limited to a particular category of predicates. In this and the following chapter it will be argued that what implicature theory identifies as Q1 implicature, should rather be viewed as a type of inference dependent on the information structure of the discourse. This type of inference is sometimes referred to as exhaustive interpretation (see e.g. Groenendijk and Stokhof (1984)). In section 3.3 I will put some flesh to this hypothesis.

The reader should keep in mind that what I am trying to do in this and the following chapter is advocate a different explanation for well-known facts. Many of the examples will look familiar to anyone who is acquainted with the Radical Pragmatics literature, but by looking at them in a different way, I hope to make the idea of a different explanation for these well-known cases plausible. But first some more preliminaries are necessary.

3.2 Formal underpinnings

In order to make the analyses presented in this and the following chapter viable it is necessary to present some formal underpinnings. We shall be brief about them, presenting the formal aspects only in merest outline, and, where possible, refer to other works where more elaborate arguments are developed. The formal context is that of Seuren’s Discourse Semantics, in tandem with his theory of Semantic Syntax, as presented in Seuren (1985) and (1996). One of the fundamental claims which this theory makes is that there exists a level of semantic analysis (SA) which provides the input for both the syntax and the semantics. The syntax relates SA and surface structure (SS) with each other. The semantics is a form of representational semantics, in which a semantic discourse representation (called Discourse Domain or DD) is formed by a process of incrementation. In this section I will discuss some aspects of the syntax and semantics of questions and answers, in so far as they are relevant for the study undertaken here.

In this and the following chapter, the SA of sentences will appear in the discussion when it is necessary to discuss their truth conditions. One of the benefits of introducing SAs is that they allow for a uniform truth condition schema for all sentences: “truth results just in case the n-tuple of the reference values of the terms is a member of the extension set of the predicate.” (1985:113) SAs are formed by rewrite rules, which define their well-formedness conditions (see 1996: 24/5).

Surface questions and their answers are assumed to be derived from the same underlying cleft structure. The SA of the question in (6a) is as represented in (6a’) in its linear form and in (6a") as a tree structure:
(6) a. Who stole the book?
   a.' :x [x stole the book] be_sp WH?

The presence of the verb be_sp as main predicate is essential for both the syntactic and the semantic treatment of this SA. To start with the semantic aspect, this so-called “specifying be” must be distinguished from the predicative be and the be stating identity. The semantic operation it represents is that of parameter setting or specifying a value, and these are different from straightforward predication (My grandfather is a minister) or the statement of identity (The morning star is the evening star). The semantics of be_sp will be discussed in detail in subsection 3.2.1. Syntactically, the verb be_sp takes two arguments; its subject term NP_1 has the form :x + Sx. The surface matrix predicate of a question, i.e. stole the book, occupies the NP_1 position with be_sp in deep structure; the answer to the question occupies the NP_2 position with be_sp.

In Semantic Syntax, there are several different types of syntactic rules, among which are the lexically induced cyclic rules. The cyclic rules that a predicate induces are indicated in angled brackets in the SA tree; optional rules are again in round brackets. Thus, the predicate be_sp can be seen to induce obligatory Object Incorporation and optional Lowering. Applying Object Incorporation (i.e., the inducing verb right-adopts its object NP (NP_2)) to the structure in (6a') leads to

(6) a.

---

43 In the syntactic treatment given here, tense is entirely left out. In Semantic Syntax, the tense routine plays an important role: not only does it take care of the syntactic tense forms, but also of the transformation from the underlying VSO-structure into the surface SVO-structure. In other words, the treatment of the tenses ensures the correct position of the syntactic subject in the surface structure.

44 The role of the operator `x` will be discussed in section 3.4.

The question in (6a) is formed from the SA (6a") by Lowering, optionally induced by the main predicate beSp: the V-cluster containing the WH-element is lowered into the NP1 subclause, replacing the variable \( x \). The verb beSp, as well as all remaining higher structure is deleted, and the subclause becomes the main clause. The WH-element is replaced by the appropriate interrogative pronoun, in this case who. Non-application of Lowering results in Who was it that stole the book? or Who stole the book was who? (cf. French Qui est-ce qui a volé le livre?).

When a question is answered, its answer appears in the SA as the second term (NP2) of beSp. E.g., (6b), as the answer to question (6a), has the SA in (6b', b"):

(6) b. David stole the book.
   b'. \( :x \) [\( x \) stole the book] beSp NP[David]

This SA can result in several different surface structures, due to the fact that the application of Lowering is optional. Lowering results in DAVID stole the book, with heavy accent on David. If Lowering does not apply, then the SA in (6b') can result in either It was David who stole the book (cleft, via it-insertion) or Who stole the book was David (pseudo-cleft). The elliptical answer David is derived from the same SA, by applying the postcyclic rule of ellipsis.

The main strength of this analysis of questions and answers is that it results in a uniform relation between the information structure of the utterances and their syntactic treatment: the topic-expression invariably occupies the NP1 position with beSp, the comment-expression the NP2 position.

### 3.2.1 The semantics of specifying be

A central element in the analysis given above is the semantics of beSp. This is not an easy matter, since beSp does not simply assign a property to an entity or an n-tuple of entities: when I say my phone number is 3616028 I do not assign the property of (being) 3616028 to the entity which is my phone number. What we have, rather, is the assignment of a value (3616028) to an entity (me) on a function or parameter (phone number). Likewise for my name is Rose and many other cases. Unfortunately, this particular class of cases has been studied very little in the existing literature. We shall here make a first attempt at providing a unified semantics for beSp; this will be done by subsequently discussing three different types of examples, which range from very obvious cases of specification to less obvious cases.
Giving a uniform semantics for these three different categories of examples will be seen to require quite an abstract characterization.

In (7)-(9), the three different categories of examples are presented. The claim is that all of these examples involve the semantic operation of specifying a value, and that their semantics essentially involves the abstract predicate $besp$:

(7)  
- a. He weighs 81 kilos.  
- b. The ship measures 44 metres.  
- c. My cousin is called John.  
- d. The planets number 9.  
- e. The ship draws two metres.  
- f. This book costs/is 39.95.

(8)  
- a. My phone number is 3616028.  
- b. My name is Rose.  
- c. The number of planets is 9.

(9)  
- a. Q <Who left?>  
  A BILL left.  
- b. Q <Who saw Mary?>  
  A TOM saw Mary.

The semantics of $besp$ is defined as follows:

**definition**  
$[[besp]] = \{<a,b>|a \text{ is an argument for a function } f, b \text{ is the value in } f \text{ for the argument } a\}$

$besp$ thus requires for truth that there be two individual objects $a$ and $b$, such that $b$ is the value for $a$ in a function $f$. It is characteristic of the first category of examples, given in (8), that this function is given by the matrix verb of the sentence. The SA underlying (8a) is taken to be as follows:

(10)  
- a. He weighs 81 kilos.

---

46 Other phrases indicating measurement are: The room is 3 by 5 metres; He stands 1.70 metres in his socks/with his shoes on.
The function required by \( be_{sp} \) can be represented as follows:

\[
\begin{align*}
\text{a} & \quad \text{b} \\
\text{object} & \quad \text{with} \\
\{ & \quad < \text{he}, 81 \text{ kg.} > \\
& \quad < \text{she}, 57 \text{ kg.} > \\
& \quad < \text{my cat}, 4 \text{ kg.} > \\
& \quad < \text{this, 8 \text{ kg.}} > \\
& \quad < \text{box} > \\
& \quad < \ldots, \ldots > \\
\end{align*}
\]

Thus we see that in this type of example, the argument \( a \) is specified in the subject position in the S directly dominated by NP\(_1\), and the value is specified in the NP\(_2\). We could call \textit{weigh}, and the other verbs given in (7), a quasi-copula.\(^{47}\) As has been said, the literature on the semantics of this type of verb is extremely scarce.

The second category of examples involves the specification of a value as well, although the name of the function appears differently in the surface structure of these sentences. The SA of (8a) is taken to be

\[(11) \ a. \ My \ phone \ number \ is \ 3616028.\]

In this type of example, both the function and its argument are specified under the S directly dominated by the NP\(_1\), as in (10). The value for the function is specified under NP\(_2\), likewise as in (10). The function is as follows:

\[
\begin{align*}
\text{a} & \quad \text{b} \\
\text{person} & \quad \text{with} \\
\{ & \quad < \text{I}, 3616028 > \\
& \quad < \text{John}, 4621890 > \\
& \quad < \ldots, \ldots > \\
& \quad < \ldots, \ldots > \\
\end{align*}
\]

\(^{47}\) The listing given in (7) is by no means intended as exhaustive.
In the third and final category, exemplified in (9), the situation is somewhat different as regards both the SA and the surface structure. The SA-structure underlying (the answers of) (9a) is taken to be

(12) a. Q. <Who left?>
    A BILL left.

In the tree structure of (12), the argument a is missing, due to the fact that tense is not expressed there. In the analysis presented here, the argument of the function is considered to be the *discourse-determined situation*. Thus, in (12) we have the function *leave*, generalized over situations, which takes as argument a specific situation and gives as value the name of a person. The missing argument can be expressed if tense is taken into account. The tense of the utterance, as incorporated in the tense operators, can be taken partly to determine the situation. In this particular type of utterance, the tense operators are located directly under the \(NP_1\)-node, as shown in (12a'):

For the precise syntactic details of this formal treatment, see Seuren (1996), especially pp. 315-21, where it is shown, in the course of a discussion of the syntactic aspects of (pseudo)clefting, that the tenses are most appropriately assigned this position in the SA-structure.
Apart from the tense operators, the sentence may contain further specifications which add to the determination of the discourse-determined situation, such as explicit indications of time and place. And, finally, the wider context, both linguistic and nonlinguistic, may serve to specify the discourse-determined situation. Thus, if the exchange in (9a) occurs in the course of a conversation about John's party, which was last night, then this is part of the specification of the discourse-determined situation. The function and its argument, as contained in the S directly dominated by the NP₁ may be paraphrased as the leaver of the situation or the situation’s leaver. Schematically, this is the function:

\[
\begin{array}{c|c|c}
   & a & b \\
   \text{discourse-determined situation} & \text{the one who leaves} \\
\hline
   \langle & S₁ & \ldots \rangle \\
   \langle & S₂ & \text{Bill} \rangle \\
   \langle & S₃ & \ldots \rangle \\
   \langle & \ldots & \ldots \rangle \\
\end{array}
\]

The same type of analysis is given for the exchange in (9b), which involves a more complicated function than (9a):
In the examples (9a) and (b), with the analyses (12) and (13), the implicit question given is essential for obtaining the specification reading: "the person who left was Bill" and "the person who saw Mary was Tom". Obviously, both Bill left and Tom saw Mary can occur in other contexts, receiving other readings.

It is crucial to the semantics of besp that it gives the value of a function for an argument, thus mapping every argument to a single value. The consistent property of all three categories discussed above is the fact that the specification asked for, the answer to the question, the value of the function for a specific argument, occupies the NP2-position with the abstract predicate besp. What varies is the way in which argument and function are given.

### 3.2.2 Relative informativeness

There is one notion, very central to the theory of Radical Pragmatics, which will receive a new definition here, the notion of relative informativeness. As was discussed earlier in section 2.1, according to implicature theory the exclusion inference depends for its derivation essentially on the notion that the speaker could have made a different, semantically stronger or more informative, statement than he actually did make. E.g., if a speaker utters (14a),

(14) a. Some of my pupils failed the test.
   b. All of my pupils failed the test.
then, because he failed to use the stronger alternative (14b), this is what is negated in the
implicature: "(as far as the speaker knows) not all of his pupils failed the test". Thus, for any
utterance, the assessment which (if any) are the stronger, more informative alternatives plays
a crucial role. As we discussed in chapter 2, this relationship of relative semantic strength was
commonly defined in terms of one-sided entailment ($P$ is semantically stronger than $Q$ iff $P$
entails $Q$ but $Q$ does not entail $P$), but notably Hirschberg (1991) signalled that a more liberal
notion of semantic strength is needed. However, what these approaches have in common is
that they attempt to define a hierarchy of informativeness among sentences (again in
isolation), that is, among the statement actually made and any available alternatives.

In this study a different notion of relative informativeness will be employed, where
entailment plays a role, but only relative to a discourse. Informativeness is first and foremost a
discourse notion; a sentence by itself is not informative. An utterance is only informative
relative to its context, where the context is not limited to the (implicit) question to which the
utterance functions as an answer, but comprises all of the previous discourse. Only
secondarily is it possible to compare levels of informativity of different utterances, only, that
is, if they are set against the same context.

We define informativeness in terms of the valuation space of an utterance, i.e. the set of
possible situations in which it is true\(^{48}\). In the definition, the valuation space of $P$ is
symbolized as $\mathcal{V}$, the result of updating a context $C$ with $P$ is written as $C + P$. Model-
theoretically, updating $C$ with $P$ has the form of intersecting $\mathcal{V}$ and $\mathcal{V}'$:

\[\mathcal{V}' = \mathcal{V} \cap (C + P)\]

The condition that $\mathcal{V}' \neq \varnothing$ expresses the condition that in any discourse inconsistencies
should be avoided. The condition that $\mathcal{V}'$ be a proper subset of $\mathcal{V}$ expresses the condition
that the context, after updating with $P$, must be true of a smaller set of situations than $C$
without $P$. If, after the addition of $P$, the set of situations in which $C$ is true, i.e. $\mathcal{V}$, is
unchanged, $P$ is not informative in $C$. (If $\mathcal{V}'$ is larger than $\mathcal{V}$, $P$ implies a correction on
$C$, by taking away some of the information contained in $C$.

If one wants to calculate relative informativeness of two utterances $P$ and $Q$ in this
system we thus have a three-way relationship among these two utterances and their common
context $C$. Assuming that both $P$ and $Q$ are informative in $C$, in the sense defined above, the
assessment of relative informativeness between $P$ and $Q$ thus amounts to a comparison
between the sets $\mathcal{V}'$ and $\mathcal{V}'$. Ideally, one would wish to say that "$P$ is more
informative than $Q$ in $C$ iff both $P$ and $Q$ are informative in $C$ and the cardinality of $\mathcal{V}'$ is
smaller than the cardinality of $\mathcal{V}'$." But this is not practical since it must be assumed that
any set $\mathcal{V}'$ is infinitely large, which makes a comparison of cardinalities senseless.

Then, although one might intuitively think of relative informativeness also with respect
to two disjoint answers to a context question, as when in the context <Who killed Mary?> the
two answers A woman did and Eric did are compared, our definition of relative
informativeness is not meant to cover such cases. What we are interested in is the class of
cases where $P$ and $Q$ are not semantically or logically independent, but $\mathcal{V}'$ includes $\mathcal{V}'$ or
vice versa, that is, where $\models Q$ or $Q \models P$, and hence $C + P \models C + Q$ or $C + Q \models C + P$. We
therefore give the following definition:

---

48 This notion is due to Van Fraassen (1971).
**definition**

$P$ is more informative than $Q$ in $C$ if both $P$ and $Q$ are informative in $C$ and

$\langle C + Q \rangle \overset{\text{def}}{\subseteq} \langle C + P \rangle$.

The two sets $\langle C + P \rangle$ and $\langle C + Q \rangle$ each indicate a set of possible situations. If one set includes the other, the latter is the more informative of the two. This definition captures the kind of relative informativeness that is relevant here. It adequately characterizes the class of situations in which there is a relation of one-sided entailment between the utterances to be compared:

(15) context \text{<How did Pete do on his exams?>}
   
   a. Pete failed two tests.
   
   b. Pete failed maths and biology.

In this example, (15b) entails (15a) but not vice versa, and (15b) is more informative than (15a). There are also contexts where, though $P$ entails $Q$, $Q$ is not informative at all, e.g. if the context already entails $Q$. An example is:

(16) context \text{<Why doesn't John's widow get the money of his life insurance policy?>}
   
   a. John is dead.
   
   b. John committed suicide.

(16b) entails (16a), but not vice versa. Nevertheless, (16b) is informative (it is a common condition in life insurance policies that the deceased should not have taken his or her own life), but (16a) is not, as it is already implied by the context. Thus, in the case of (16) we cannot really say, according to our definition, that (16b) is more informative than (16a).

The notion of relative informativeness of an utterance with respect to a context can be used to define a sufficient condition for pragmatic well-formedness for discourse:

**pragmatic principle**

A discourse is well-formed if every successive utterance is informative with respect to its context.

Thus, the following discourse is well-formed:

   a. Pete failed two tests.
      <Which ones did he fail?>
   
   b. He failed maths and biology.

However, reversing the order of the utterances does not result in a well-formed discourse:

(17) \text{<How did Pete do on his exams?>}
   
   a. Pete failed maths and biology.
      ?<How many tests did he fail?>
   
   b. ?He failed two tests.

The uttering of (17b) is not informative in the given context, because, informally stated, it does not add any new information. In formal terms, $\langle C + (17a) \rangle + (17b)$ is a subset of $\langle C + (17a) \rangle$, but not a proper subset.
3.3 Exclusion inferences and exhaustive interpretation

Exhaustivity is a property of the interpretation of utterances in general. Its workings can informally be described as follows: some aspects of an utterance indicate that it has to be interpreted exhaustively. When the hearer perceives these clues correctly he will draw the conclusion that the utterance covers some specific (set of) objects/situations in the world completely (it exhausts the objects/situations described). E.g., when an utterance talks of five naughty pupils, then on an exhaustive interpretation the hearer will assume that there are only five naughty pupils and not more, that there are exactly five naughty pupils. Intuitively, exhaustivity has to do with "speaking the whole truth", not with "speaking the truth" simpliciter.

The intuitive similarity between exhaustive interpretation and the exclusion inference appears when the description above is compared with Levinson's concluding remark of his discussion of Q1 implicature:

in many cases the implicatures can be glossed by adding only to the propositional content of the sentence, e.g. Nigel has only 14 children, the flag is only white, Harry only got a fine. (1983:106/7)

The informal characterisation of exhaustive interpretation allows us to determine two of its properties. First of all, it pertains only to "some specific (set of) objects/situations". This formulation indicates that it is not necessarily the whole utterance that is interpreted exhaustively - it may also involve only a part of it. Consequently, we need a method to determine which part of the utterance it is that is interpreted exhaustively. In the view developed here, the determining factor is the context, specifically the question to which the utterance under consideration is an answer. It is the comment which is interpreted exhaustively. The (set of) objects/situations covered completely form the denotation of the lexical items in comment-position.

Secondly, the process of exhaustive interpretation has truth-conditional results. E.g., in (18)

(18) Q How many pupils are there in your class?
    A 31

the answer, specifying the number of pupils in the speaker's class, is to be interpreted exhaustively, i.e. as saying "exactly 31". If it turns out that there are 33 pupils in the speaker's class, then he will not have spoken truthfully. A similar, but not identical, example, not involving a numeral, is (19)

(19) Q What is your profession?
    A I am an architect

---

49 Two of these aspects are information structure (topic-comment modulation) and intonation. The specific role of intonation, and its possible connections with information structure, are not the topic of the present study. For a discussion, see Van Kuppevelt (1991).

Here again, if it turns out that the answerer is also a practising doctor, then he may not have spoken the whole truth, depending on relevance criteria in the situation at hand. The difference with (18) consists in the fact that the presupposition of (19) that the addressee has just one profession may be false; since it is not unheard of that a person has two, or even more, professions. The presupposition of (18), however, i.e. that the set of pupils in the class has just one cardinality, cannot be false: any finite denumerable set has one and only one cardinality. If what is asked for is that cardinality then only one answer can be true. The answer in (19), on the other hand, may well be called true but it is incomplete, and for the complete answer the presupposition implicit in the question has to be corrected. In section 4.3.2, corrections like these will be discussed.

This basic intuition leads to the prediction that any continuation on an exhaustively interpreted phrase will be a correction, which has to be marked in some ways (lexically, intonationally). On the other hand, modifications to phrases that are not interpreted exhaustively do not need any special markings. Here is an example of correction:

(18)' Q  How many pupils are there in your class?
    A  31. No wait, 33.

The information structure of this piece of discourse is as follows:

(18)'' Q1  How many pupils are there in your class?
    A1  31
    <Q2 Are you sure?>
    A2  No wait, 33.

In this view, correction involves replacing the initial answer to some question by a different answer given on closer scrutiny. This process requires explicit marking (lexically, intonationally), as it is not the default procedure in an ongoing discourse. When a modification of material not in comment-position (thus not interpreted exhaustively) is necessary, then this needs much less emphasis:

(20) Q  How many sheep do John and Jane own?
    A  Four. Besides, /In fact, /But her name is Jean.

This way of looking at exhaustive interpretation of lexical material in comment-position provides a unitary view for something that implicature theory is forced to explain asymmetrically. In example (3'), the correction involved a higher number than the initial answer. In implicature theory, this would be a case of implicature cancellation. That correction is not the same process as implicature cancellation will be discussed in more detail in the next section. However, at this point we can already point out one difference between the two notions. Implicature cancellation always involves naming a stronger value on the same scale than the one initially mentioned, in the case of numbers a higher number. When the continuation involves a lower number, as in (20),

(21) Q  How many pupils are there in your class?
    A  31. No wait, 29.
Implicature theory would presumably regard this as an example of correction, just as I do. But then we have an asymmetrical explanation for two examples which are similar in all other respects. According to implicature theory, when the continuation involves a stronger value (e.g., a higher number) it is an instance of implicature cancellation, but when it involves a weaker value (a lower number) it is an instance of correction. This asymmetry is not supported by any empirical differences.

3.4 The formal semantics of numerals

In Radical Pragmatics, the choice for the "at least" reading as the fundamental, semantic reading for numerals is not motivated other than by scientific elegance. The source of this view is the theory of cardinals developed in formal semantics, more specifically the Montague tradition. Here, cardinals are quantifiers, and their definition is based on the definition of the existential quantifier. Even so, however, there is no generally accepted, standard analysis of cardinal numbers. I will present one instance of this approach, taken from a textbook for students (Cann 1993: chapter 6).

Cann discusses both the Russellian quantificational view of numbers and an approach based on generalised quantifiers. His exposition is based on two closely related assumptions which are found in almost all theories on the semantics and pragmatics of cardinal numbers in natural language. The first one he shares with Radical Pragmatics: the truth-conditional semantic value of any numeral $n$ is "at least $n". The second assumption is that cardinal numbers in natural language function as quantifiers.

Classical quantification theory has two quantifiers, the universal and the existential one. Any other quantifiers will have to be represented in terms of one or both of these two, which, as is well-known, is not always possible (e.g. most cannot be rendered in Russellian quantification theory). For reasons of exposition, Cann contrasts (at least) two and both with each other, and for the moment I will follow him in this. The semantics for both that he assumes is "exactly two". Cann gives the formulas in (22b) and (22d) as (according to his assumptions, truth-conditionally correct) translations of the sentences in (22a) and (22c).

(22) a. At least two students laughed.
   b. $\exists x \ [\exists y \ [\text{student'}(x) \ & \ \text{student'}(y) \ & \ (x=y) \ & \ \text{laugh'}(x) \ & \ \text{laugh'}(y)]]$
   c. Both students laughed.
   d. $\exists x \ [\exists y \ [(\exists z \ [\text{student'}(x) \ & \ \text{student'}(y) \ & \ (x=y)]] \times [(z=y) \ & \ (z=x)] \ & \ \text{laugh'}(x) \ & \ \text{laugh'}(y)]]$

Cann rejects these translations because they are not adequate representations of the linguistic objects: "[they] are very complex and counter-intuitive". Moreover, they show a weak point of classical quantification theory: they introduce "sentential connectives that are just not represented in the English syntax" (1993:187/8).

Cann subsequently presents an alternative analysis in terms of generalised quantifiers. Within this theory, it is possible to formulate the satisfaction conditions of many different quantifiers directly, thus not in terms of the universal or the existential quantifier, and quantifiers like most or half are, moreover, expressible without difficulty. In the case of (at least) two and both the satisfaction conditions become:
Both of these quantifiers put constraints on the cardinality of the intersection of two sets, that denoted by the common noun and that denoted by the VP, but the difference between them is that two' requires the intersection of these two sets to have two or more elements while both' requires the cardinality of the set to be exactly two. Under the analysis of the interpretation of the NP two students in [23a],

\[
\text{[23] a. } [\text{two'}(\text{student'})]^{M,g} = \{X \in A \mid |X \cap \text{[student']}^{M,g}| \geq 2\}
\]
\[
\text{b. } [\text{both'}(\text{student'})]^{M,g} = \{X \in A \mid |X \cap \text{[student']}^{M,g}| = 2\},
\]

where \(A\) is the set of all individuals in the model

the formula translating Two students laughed is true just in case the intersection of the extension of student' with that of the VP, laugh', has two or more members (i.e. if the set of students who laugh has a cardinality of greater than or equal than two). (1993:189).

Cann's approach has several weaknesses; I will first discuss two aspects concerning the truth-conditions that Cann assumes for (at least) two and both, and then go on to the more fundamental assumptions mentioned above.

First, one of the (undesirable) consequences of this analysis, as Cann implicitly acknowledges, is the fact that at least two and two come out as (semantically, truth-conditionally) synonymous: Cann speaks of "the treatment of the quantifier (at least) two'. That this synonymity is incorrect is easily shown. In a context such as (24b),

(24) How many books did Peter bring along?
   a. Two.
   b. At least two.
   c. Two or four.

(25) He brought along three books.
   a. Two.
   b. At least two.
   c. Two or four.

a report given as in (25) could be true, as at least two means "two or more". However, in a situation as in (24a), an utterance of (25) would inevitably be false. And finally, in the case of (24c), the "at least" reading for two (or four, for that matter) would lead to straightforward absurdity. The context in these examples, specifically the question to which (24a-c) are answers, is essential.

Secondly, the analysis of both as a quantifier is simply wrong. Both presupposes a set with cardinality two and states that the predicate asserts something with respect to all members of this set; it is a dual\(^{51}\) anaphor, and its meaning is "the two". If the situation were as Cann presumes, then the uttering of both in the second part of (26a) would have to be superfluous, as it merely repeats the statement made by two in the first part of (26a).

(26) a. The club had two members, John and Peter. Both were living in High Street.

\(^{51}\) There are languages which have, next to the grammatical categories of singular and plural, a third category named dual, which refers to two items. These languages have, next to a singular and a plural, a dual form for nouns, and often also separate articles for each of these three numbers. In English, duality is not a grammatical feature, but it appears as a lexical feature in both.
b. John and Mary entered the room. Both were laughing.

However, this piece of discourse is perfectly acceptable. Both can be seen to function as a definite anaphor, referring back to a set with two members which was previously introduced. It can refer anaphorically as well when two is not explicitly present, as in (26b). Consequently, Cann's analysis of both as a quantifier meaning "exactly two" is not an adequate account of the semantics and discourse function of both.

Now back to the two assumptions mentioned at the beginning of this section: Cann regards any numeral \( n \) as a quantifier with the truth-conditional meaning "at least \( n \)". This view on the semantics of numerals is the expression, in semantic terms, of the judgement that in all situations where the sentence \( X \) has three members is true, the sentence \( X \) has two members is also true (entailment). From this it is concluded that, conversely, if \( X \) has two members, then the possibility that \( X \) has three members cannot be excluded. Therefore, the semantics of a numeral \( n \) is "at least \( n \)".

In Russellian quantification theory, this is easily translated by the use of the existential quantifier, and so the semantics of a numeral \( n \) becomes (in abbreviated form) \( \exists nX \). The semantics of numerals in generalised quantification theory looks more adequate as a semantics for natural language, but it gives essentially the same satisfaction conditions as classical quantification theory. Two is a version of the existential quantifier ("XIIA"), and its satisfaction condition is "at least \( n \)" ("\( \geqslant n \)").

This theory is in itself coherent, but runs into one problem when asked to account for the use of numerals in natural language, as I will show in detail in the next section. There are situations where, when it is stated that \( X \) has \( n \) members, the possibility that \( X \) has \( n+1 \) members is in fact excluded. These are the situations that implicature theory wants to cover by a generalized Q1 implicature. But there is another way, and that is to treat the numeral not as a quantifier, but as a predicate specifying a specific value (in the cases under discussion here, a cardinality).

### 3.5 Exhaustive and non-exhaustive interpretation of numerals

In this section I will study the interpretation of numerals in context, where the context is represented in the form of a question. A question defines a topic/comment-modulation for the utterance which functions as its answer. This information structure influences the interpretation of the numerals in these utterances, as was observed in Seuren (1993). When a numeral is (part of) the comment-expression of the utterance, the “exactly” interpretation is truth-conditional, the possibility that a higher number obtains is excluded. The prototypical example of this situation is when numerals appear in answers to a “how many” question.

---

\(52\) As said in the beginning of this section, within the tradition of formal semantics there is no generally accepted analysis of cardinal numbers. Cann considers two and at least two as synonymous, but e.g. Gamut (1991, vol. 1: 104/5) considers two and exactly two as synonymous. This can be concluded from the fact that they present the formulae and their translations in (i) as "sentences expressing the numerals one, two and three."

(i)

There is exactly one \( x \) such that \( Ax \)
\[ \exists x \forall y [Ay \times x=y] \]
There are exactly two \( x \) such that \( Ax \)
\[ \exists x \exists y [x\neq y \& \forall z [Az \times [z=x \& z=y]]] \]
There are exactly three \( x \) such that \( Ax \)
\[ \exists x \exists y \exists z [x\neq y \& x\neq z \& y\neq z \& \forall w [Aw \times [w=x \& w=y \& w=z]]] \]
When a numeral is part of the topic-expression of the utterance, the "exactly" interpretation has a different status. I will first look at structures with the numeral in comment-position, then at structures with the numeral not in comment-position.

One of the chief arguments in this and the following chapter will be the intuitive judgement that many of the alleged examples of implicature cancellation that can be found throughout the Radical Pragmatics literature are, at least when viewed in isolation, unacceptable or odd (see section 2.3.2). An attempted explanation for this intuition, based on a different view of the nature of the exclusion inference, will be given in section 4.3.5.

### 3.5.1 Numerals in comment-position

With respect to numerals in comment-position there are in theory two possibilities: either the numeral exhausts the comment-expression, or it is only part of it. Although this difference is not reflected in the interpretation, for ease of exposition I will start with the first category. The prototypical case is the use of a numeral in an answer to a "how many"-question.

(27) Q  How many sheep does John own?  
   A  He owns **four** sheep.

(28) Q.  How much did you pay for that camera?  
   A.  I paid **300** guilders.

Example (28), though not literally a "how many"-question, can easily be paraphrased as one: *How many guilders did you pay for that camera?* In “how many” questions, the question asks for a specification of the cardinality of some set, in (27) the set of sheep owned by John. The (unspecified) cardinality of the set of sheep owned by John is topic. In (28), the specification asked for concerns an amount of money, the cardinality of the set of guilders payed for a particular camera. The specification asked for is given by the numeral, which is therefore in comment-position. When a numeral is used to name (specify) a cardinality in this way, it is interpreted exhaustively. Continuation with a higher number, unless explicitly marked as being a correction, leads to an unacceptable piece of discourse.

(29) Q1  A  How many sheep does John own?  
   A1  B  He owns **four** sheep, *in fact ten.*

In answer to (Q1), (A1) will give rise to surprise on the part of speaker A. In the view developed here, speaker B is in fact giving two contrary answers to one question. Speaker A could voice his surprise by asking speaker B which of the answers he gave is the one he wants to stick by:

(29)' Q2  A  How many sheep does John own, four or ten?  

Speaker A's surprise would not have arisen if B had uttered a correction, as in

(29)" Q1  A  How many sheep does John own?  
   A1  B  He owns **four** sheep.  
   <Q2  Are you sure?>
A2  B  Sorry, I'm wrong, he owns ten.

(28) Q1.  A  How much did you pay for that camera?
   <Q2. Are you sure?>

Continuation with a higher number counts as implicature cancellation in implicature theory; in the configuration discussed here, it does not lead to an acceptable piece of discourse. Continuation with a lower number shows a similar pattern: it is only acceptable when it is adequately marked as being a correction.

(30) Q1.  A  How many sheep does John own?
   Q2.  A  What do you mean, does he own three or four?

Both (29) and (30) are illustrations of the fact that under an exhaustive interpretation, when indicating a cardinality, the entailment relations often judged to exist among sentences containing numerals do not hold. Or, in other words, when numerals indicate a cardinality there is no implicational ordering.

This is the case not only for examples such as (27) and (28), where the numeral exhausts the comment, but also for examples where the numeral is only part of the comment. The unifying characteristic is the fact that the numeral is used to indicate a cardinality.

(31) Q.  What did John do?
   A.  He bought four books.

Again in this category, the numerals will be interpreted exhaustively. Continuation with a higher number (implicature cancellation) is only acceptable if duly marked, which signals that we are dealing with a correction here.

(32) Q1.  What did you do?

One may object that, even when a numeral appears as the answer to a "how many" question, the "exactly" reading is not always truth-conditional, especially when higher numbers are at issue. E.g., in (28) above, the price of the camera may have been £349, and not £350. Should we say the speaker was untruthful? The common sense answer to this question would be no, and I think this is the correct theoretical answer as well. In daily life, when quantities become larger, numbers are used more as an approximation than as indicating an exact quantity. This has nothing to do with the system of language, which is our topic here, but with human skills and habits in handling numbers. Why people tend to use approximate rather than exact numbers for higher quantities, and how they do this, probably is a psychological issue; some brief remarks may suffice here. In general, people are seen to deal with smaller quantities more easily and more exactly than with larger quantities, whether it be the population of a city, sums of money, weights, distances or temperatures. (On the other side of the scale extremely small quantities, i.e. between 0 and 1, provide difficulties as well.) Secondly, there are large differences between individuals with respect to their skills in
handling either very large or very small numbers; it is something one can learn. And thirdly, the precision with which quantities are determined varies from situation to situation; a single gram of some substance may be very important in science (e.g. when preparing a medicinal drug), but not in daily life (e.g. when baking a cake).

The upshot of these brief remarks is that the semantics and pragmatics of numerals developed here will not suffice to explain all uses of numerals. The context is generally much wider (non-linguistic, psychological) than the single question which I use to represent it. Nevertheless, this single question allows us to draw some substantial conclusions on the influence of information structure on the interpretation of numerals.

The question now is, what should we assume to be the semantics of numerals, on the basis of the examples discussed in this section? It is clear that the assumption that any numeral \( n \) means “at least \( n \)” is incorrect. The origins of this view were discussed in section 3.3, and there we find the starting point for a different view of the semantics of numerals. In all the examples discussed in this section, numerals were used to indicate the cardinality of a set. This is a definite cardinal number, not an infinite disjunction of numbers (“at least \( n \)” = “\( n \) or \( n+1 \) or \( n+2 \) or…ad inf.”) or an unspecified number of which only the lower bound is given (\(|X| \geq n\)). There is a conventional notation for this:

\[
(33) \quad |\{x, y, z\}| = 3,
\]
or in general

\[
(34) \quad |X| = n, \text{ where } X \text{ is a set and } n \text{ is a cardinal number}
\]

Note that this makes numerals higher order one-place predicates, whose term refers to a set. So by definition,

\[
\text{definition}
\]

\[
[[\text{num}_n]] = \{X \mid |X| = n\}
\]

where \( \text{num}_n \) ranges over numerals \( n \) and \( X \) ranges over sets. 

\( \text{E.g., } [[\text{four}]] = \{X \mid |X| = 4\} \)

I will adopt this definition as a unitary semantics for numerals, and show that it makes it possible to account adequately for various uses of numerals. It is possible to incorporate this view of the semantics of numerals into various semantic theories. I will further explicate the semantic notions involved within the context of Seuren’s Discourse Semantics (1985) as introduced in section 3.2. Due to the overall structure of this theory, various aspects of SAs are motivated by syntactic considerations - I will mention these aspects but I will not discuss the considerations behind them. Details of the process of incrementation can be found in Seuren (1985, 1994).\(^{53}\)

\(^{53}\) Other views on the semantics of numerals within the context of a form of representational semantics can be found in Kadmon (1987: chapter 2) and Kamp and Reyle (1993: chapter 4). Kadmon adopts a semantics of numerals akin to that of formal semantics, thus where \( \text{three} \) is truth-conditionally equivalent to “at least three.” She furthermore appeals to the theory of implicatures in order to alter the truth-conditional meaning of the numeral into the correct interpretation. Kamp and Reyle do not discuss the semantics of numerals as a separate issue, but numerals do appear in their extensive discussion of the semantics of the plural. They adopt a semantics similar to the one adopted here: \( \text{three} \) and \( \text{exactly three} \) are equivalent, both are represented in the form of a condition \(|X| = 3\). However, no motivation for this choice is given.
When a numeral is in comment-position, question and answer are presumed to have the following semantic structure (informally rendered):

(35) Q  How many sheep does John have?
       \[the cardinality of the set of sheep that John owns\] is WH?
A  He has four sheep.
       \[the cardinality of the set of sheep that John owns\] is four.

The semantic process involved is parameter setting or value assignment. The cardinality of a set is a parameter, the numeral assigns it a certain value. This is the SA of (35A):

(35)'

\[
\begin{array}{c}
V_1 \quad \text{be-sp} \\
<\text{OI,L}> \\
S_0 \\
NP_1 \\
NP_2 \\
\vdots \\
V_2 \quad \text{with cardinality} \\
\lambda z \quad \text{function} <\text{OI,L}> \\
S_1 \\
\vdots \\
V_3 \\
\vdots \\
V_4 \quad \text{NP} \\
\vdots \\
\vdots \\
\vdots \\
\vdots \\
\vdots \\
\vdots \\
\text{NP}_x \\
\text{value} \\
\end{array}
\]

The \(\lambda\)-operator yields a predicate \((V_3)\) denoting the intersection of its argument sets. Then, \([[V_3]]\) is the set of all sheep that John owns (the intersection of [[S_3]] and [[S_4]]). \(NP(y)\), being a variable, rotates over the set of individuals in the model, yielding again the set of sheep that John owns as [[S_2]]. The operator \(\text{\`y}\) has a purely syntactic function, turning \(S_2\) into an NP, \(NP_3\), the subject of \(V_2\). \(NP_1\) now says "the \(x\) such that the set of sheep that John owns has the cardinality \(x\)". The predicate \(\text{be-sp}\) now checks if the specification given in the \(NP_2\)-position is correct, i.e. whether [[NP_3]] has the cardinality "four" (for, by definition, [[four]] = \(\{X | |X| = 4\}\)). If this specification is correct, \(S_0\) is true, if not then it is false. This is the exhaustive reading, or, as I will sometimes say, the cardinality reading. This reading is opposed to the existential reading, which is commonly referred to as the "at least" reading.

The syntactic treatment (according to Seuren 1996) may, in principle at least, be taken to proceed according to the following steps. The first cyclic application of \(<\text{OI,L}>\) (Object Incorporation and Lowering) gives "own John sheep". The second cyclic application of \(<\text{OI,L}>\) gives "own John sheep with cardinality \(x\)" or "own John \(x\) sheep", where \(x\) stands for an expression denoting a cardinality. The third application of \(<\text{OI,L}>\) gives "own John FOUR sheep", with nuclear predicate accent on FOUR. Standard tense treatment then gives the required NP-VP-structure "John owns FOUR sheep". There are, of course, many details of this syntactic treatment that cannot be discussed here. Nor can it be claimed that all syntactic problems for this class of constructions have been solved. All that can be given here is a general and global indication of how the relation between the SA-structures and the surface
structures concerned may be established by means of a parsimonious and well-constrained set of transformational rules.

The entailment relation among sentences containing numerals that formal semantics intends to account for (see the discussion in section 3.3), is misunderstood. This relation among sentences does not constitute a fact about the semantics of numerals, but it follows from a property of sets: if a particular set X has four members, then it has four subsets with each three members, and six subsets with each two members, and again four subsets with one member. From this set-theoretic property it follows that (36) entails (37), but only if (37) receives the existential (“at least”) reading, as explained in the following section:

(36) Barbara has four children.
(37) Barbara has three children.

However, it is false to take this property of sets as the main semantic aspect of numerals, as is standardly done in formal semantics. What the numerals do, in the cases under discussion, is predicate a specific cardinality of a set. It is sets that are being predicated over, and sets have certain properties - but these properties are independent of the linguistic semantics of numerals.

Finally, in the view adopted here, $n$ in comment position and exactly $n$ have the same satisfaction conditions. That they are, however, not necessarily pragmatically equivalent will be discussed in section 4.5.

### 3.5.2 Numerals not in comment-position

When numerals are not part of the comment expression of an utterance, we generally find the interpretation pattern described by Radical Pragmatics: the preferred interpretation of a numeral is “exactly $n$”, but this is easily overruled in favour of an “at least” reading. In this section I will show that this same pattern can be achieved on the basis of the semantics of numerals as defined in the previous section.

An argument for maintaining the semantics for numerals defined in the previous section lies in the fact that, just as with numerals in comment-positions, there is no difference, with respect to correction, between replacement by a higher or by a lower number. For a lexical item to be in non-comment position it has to be part of the topic-expression of an utterance, or be part of a feeder or topic-neutral utterance. A simple example where a numeral is part of the topic-expression is this:

(38) Q Who owns four sheep?
    A John owns four sheep.

The comment in this case is John; this is interpreted exhaustively, to say that (within the relevant domain of discourse) only John owns four sheep, nobody but John owns four sheep. However, with respect to the numeral there are two possibilities. If the number of four sheep was introduced earlier in the discourse in comment position, the exhaustive interpretation of the numeral seems to be maintained, as in (39):

(39) Q1 How many sheep does Peter have?
A1 He has four sheep.
Q2 Is there anyone else, apart from Peter, who owns FOUR sheep?
A2 John owns four sheep.

This default strategy (i.e., if a certain numeral \( n \) has received the cardinality reading once in a discourse, it will maintain that reading throughout that same discourse) can be accounted for by appealing to the pragmatic principle of well-formedness that we defined earlier: according to the definition of relative informativeness, an utterance \( P(n) \) is more informative when interpreted as "exactly \( n \)" than as "at least \( n \)". Thus, the discourse as a whole would become ill-formed if the stronger interpretation for a numeral were replaced with the weaker one in the same discourse.

On the other hand, if the number of four sheep is newly introduced into the discourse by (38Q), the interpretation that John has no more than four sheep is only the preferred interpretation. The possibility that John has more than four sheep cannot be excluded on the basis of (38A). In semantic terms: the truth conditional meaning of four in this case is “at least four”. This situation fits neatly into the schema of implicature theory: the preferred interpretation “exactly four” is the result of an implicature, which can be cancelled without a contradiction arising:

(40) Q Who owns four sheep?
   [The person such that there are 4 sheep he owns] is WH?
A John owns four sheep, in fact he owns twenty.

I will, however, propose a different analysis for this example: the numeral receives the existential reading, which accounts for the fact that in fact-correction is possible. Moreover, there is an implicit how many question between the two parts of the answer; therefore twenty receives the cardinality reading:

(40)' Q Who owns four sheep?
   [The person such that there are 4 sheep he owns] is WH?
A John owns four sheep
<Q How many sheep does John own?>
A (in fact) he owns twenty.

The technical specifications of the existential reading are as follows; it is based on the semantics of numerals as introduced in the previous section. When the numeral is in comment position (when it receives the cardinality reading), it occupies the NP\(_2\) position in the SA. To obtain the “at least” reading, the existential quantifier needs to be the highest predicate. The existential quantifier is a generalized quantifier, and its definition is straightforward and general, applying both to first- and higher order terms:

**definition**

\[
[\exists] = \{X,Y \mid X \rightarrow Y \neq \}
\]

(Since \( \exists \) does not distinguish between singular and plural, we write \( \exists_{sg} \) for the singular version, \( \exists_{pl} \) for the plural version and \( \exists_{num} \) for specific numeral versions.) A variable binding subscript \( x \) or \( y \) is standardly used with the quantifier in order to avoid ambiguity when more than one quantifier is used. The SA of John bought a sheep, where \( \exists_{sgX} \) is the singular existential quantifier, is taken to be:
The syntactic treatment is again straightforward (see Seuren 1996:300-309). The cap operator is lowered twice onto the position of the variable bound, giving "buy John `x" and "sheep `x", respectively. Object Incorporation gives "3sgx + sheep `x" for the V of S0. This V is then lowered into S1, i.e. "buy John `x", to the position of `x, giving "buy John a sheep". The tense treatment is as usual (see Seuren 1996:67-68), giving _John bought a sheep._

The terms of the generalized quantifier, NP1 and NP2, are reversed in comparison to general practice - but this does not affect its satisfaction conditions in any way. The reversal of these terms in the SA is merely determined by syntactic considerations: the standard operation of Object Incorporation requires this order. Moreover, the syntactic operator ^x reappears. In this SA, [[NP1]] = [[S1]], the set of things that John bought, and [[NP2]] = [[S2]], the set of sheep. The existential quantifier requires that the intersection of these two sets be non-empty.

In order to represent the numerals adequately, it is necessary to have a notion of plurality. Plurality involves plural power sets. The plural power set of X is defined as the power set of X minus the empty set and the singletons, or formally as

\[ P_{pl}(X) = \{ Y \mid Y \cap X \land |Y| \geq 2 \} \]

In other words, \( P_{pl}(X) \) is the power set of X such that all members of X have two or more members. Numerals combine with plural nouns, and it is possible to combine this general definition of plurality with that of the numerals. This gives the definition of the n-power set of X:

\[ P_{n}(X) = \{ Y \mid Y \cap X \land |Y| = n \} \]

E.g., if X is a set with four members \{a,b,c,d\}, then \( P_3(X) \) is by definition \( \{ Y \mid Y \cap X \land |Y| = 3 \} \) and by enumeration \{\{a,b,c\}, \{a,b,d\}, \{b,c,d\}, \{a,c,d\}\}.

This is the notion we need to account for the semantics of numerals which are not in comment-position (on the occasion of their first use in the discourse at hand). Consider (42) to be an example of a numeral in a feeder:

(42) a. John bought five sheep.

In this context-neutral interpretation, its SA is taken to be
The existential quantifier, which is the highest predicate in this SA, operates on the terms NP₁ and NP₂. The operator ∃ₓ under the node NP₁ is comparable to the operator ∧ₓ we saw in the previous section: it has the syntactic function of turning an S into an NP, but ∃ₓ also has the semantic function of forming the plural power set of the input set, in this example [[S₁]]. So [[NP₁]] is the set of sets of at least five things that John bought, or {Y| Y ∩ {x| John bought x} & |Y| ≥ 5}. NP₂ is formed in the same way as NP₁, thus [[NP₂]] = {Y| Y ∩ {x| sheep(x)} & |Y| ≥ 5}.

The existential quantifier now operates on these two terms, giving as a result that the intersection of [[NP₁]] and [[NP₂]] contains at least one member Z such that |Z| = 5. The general definition of the numeral existential quantifier is

\[ [[∃ₙₙ_x]] = \{X,Y | Ppl(X) « Ppl(Y) « [[n]] \neq \emptyset \}, \]

where X and Y range over sets and n ranges over names of cardinal numbers.

Thus, (42a) is true iff there is at least one set of sheep that John bought which has cardinality five. This explicitly leaves open the possibility that there are more such sets, so that John bought more than five sheep. This seems to be a viable truth-conditional interpretation for the quantifier analysis.

The preferred interpretation of numerals not in comment-position is different, however: the preferred interpretation is “exactly n”. In Discourse Semantics, this interpretation gets the form of a default-interpretation and is due to the rules with respect to specific reference. I will discuss this complex of rules and principles briefly, only in so far as it is relevant for my purposes here. Starting point is the semantics of the existential quantifier, which is the highest predicate in the SA of (42b) above: it is a technical predicate which gives the instruction to set up a new address (one address). A DD containing such an address is “made true by at least one entity satisfying the conditions specified in [this address]. The existence of more such entities is not ruled out, but not required for truth.” (Seuren 1985:315, my emphasis)

Now the principle responsible for turning the truth-condition “at least one entity satisfying the conditions specified” into the stronger condition “exactly one entity (etc.)” is the assumption that specific reference is preferred over non-specific reference. Specific reference (Seuren 1985:459-464) is the default situation: a term refers specifically when it stands in a relation of fixed or constant designation to some individual in the verification domain V, or in other words when there is no doubt as to which individual is designated by which term (for the duration of the discourse at hand). For each term, there is exactly one
object which satisfies its conditions. In example (42a) above, where we have a plural term, we
speak of specific reference when there is exactly one set of sheep that John bought which has
cardinality five. Non-specific reference, on the other hand, appears when the verification
domain V contains several objects which satisfy the conditions of one single term, when it is,
in other words, not possible to establish which specific individual is the referent of that term.

The preference for specific reference influences, as a default strategy, the construction of
discourse domain by the listener:

.. the listener, on hearing a sentence like I have a colleague, builds up a mental
representation of the state of affairs in which that sentence is true. In doing that he
will construct a representation in which any subsequent anaphoric expression for
that colleague will refer specifically, avoiding the necessity of overriding the
default by applying the non-specific reference procedure. (Seuren i993:233)

The interpretation of the existential quantifier thus becomes an instruction for building a
discourse domain: 'an existential quantifier requires only one matching individual in order to
be true, thus sets up only one address in the discourse domain.' For the discourse at hand, this
address is supposed to have a relation of fixed or constant designation to that individual. In
cases of plural quantification, it is sets of individuals that play the part of single individuals in
singular quantification. For (42a) this means that the plural address set up of five sheep is
supposed, by default, to stand in a relation of fixed designation to a single set of sheep that
John bought, which has cardinality 5. And because we know that it is one single set, we also
know that it cannot be a subset (for any set with, e.g., cardinality 6 has six possible subsets
with cardinality 5), and thus that its cardinality is exact or maximal. This is the desired
interpretation for numerals not in comment-position.

In chapter 1 we saw that one of the problems that induced Horn to adopt the notion of
generalized conversational implicature was the interpretation of numerals (and other
predicates as well) under negation. An isolated sentence (e.g., a feeder) such as (43)

(43) John does not have three children.

is characteristically interpreted to say that John has less than three children. This
interpretation is easily achieved in the present proposal, but only when the numeral is in non­
comment position. On the existential reading, the numeral truth-conditionally means "at least
three", and its negation is "not at least three", thus "not three, four, five, ... etc.". However,
when the numeral was introduced in comment-position, as in (44),

(44) Q1  A  How many children does John have?
      A1  B  John has three children.
      A2  C  No, John does not have THREE children,
            <Q2 Then how many does he have?>
                  he has four children.

the negation simply cancels the cardinality reading of the numeral, and continuation with
either a higher or a lower number are equally possible. Those familiar with Horn's (1985,
1989) work will notice that in (44A2), the negation would be an instance of metalinguistic
negation, i.e. a negation not affecting the truth conditions of the utterance, but some other
aspect of it, in this case the Q1 implicature (for further discussion, see section 4.3.2.2).
However, on the present account such a resort to two different uses of negation is not necessary; due to the differences in information structure between (43) and (44) the numeral *three* receives two different interpretations, and the negation is the usual truth-conditional operator.

### 3.5.3 Yes/no-questions

So far in this study we have limited ourselves to examples where the context is represented by a WH-question, following the same limitation adopted in DTT (see section 1.3.2). In this section I will widen the perspective to yes/no-questions. Most statements can be turned into a yes/no-question, by attaching a question operator as highest operator in SA:

(45)

```
S
  Que
S
```

Consequently, any information structure that a statement can possess is carried over to the yes/no-question. Based on the information structure, one can distinguish two types of yes/no-question: neutral and topicalized yes/no-questions. In this section I will first show how each of these is formed, and then show that there are significant differences between them with respect to the interpretation of their answers. Finally I will argue that the failure to distinguish these two types of yes/no-question has led to unnecessary puzzles with respect to the interpretation of numerals.

The distinction between topicalized and neutral yes/no-questions corresponds to the fact that statements can perform two functions in discourse: they can function as an answer to a WH-question or not. In the latter case, they function as feeder or as what we have called a topic-neutral utterance (see section 1.3.2, but for the purposes of this section we need not distinguish these). It is characteristic of answers to WH-questions that a specific part of the statement is in comment-position (in SA: occupies the NP2 position with the abstract predicate *be*), as in (46a). If such a statement is questioned, the resulting question is a topicalized yes/no-question, as exemplified in (46b) and (c):

(46) a.  

```markdown
<Who took your books?>  
John took my books.
```

b.  

```markdown
Did JOHN take your books?/ Is it JOHN who took your books?
```
A topicalized yes/no-question can be signalled by the use of intonation, notably by heavy accent on the phrase in comment position. On the other hand, questioning a statement not functioning as answer to a WH-question leads to a neutral yes/no-question. Their semantic structure is different from that of topicalized yes/no-questions: notably, the abstract predicate $be_{sp}$ is absent:

\[(47)\]  
\[a. \text{ John took my books.} \]
\[b. \text{ Did John take your books?} \]
\[c. \]

\[
\begin{array}{c}
\text{S} \\
\text{Que} \\
\text{S} \\
\text{V} \\
\text{be-sp} \\
\text{<O1(0)>} \\
\text{S''} \\
\text{V} \\
\text{NP} \\
\text{S'} \\
\text{V} \\
\text{NP} \\
\text{S''} \\
\text{V} \\
\text{x} \\
\text{NP} \\
\text{take} \\
\text{NP} \\
\text{NP} \\
\text{your} \\
\text{books} \\
\text{John} \\
\end{array}
\]

tense operators

In section 1.4 we saw that it is characteristic of WH-questions that their answers will be interpreted exhaustively. This exhaustive interpretation can be paraphrased adequately by a cleft or pseudo-cleft construction employing the phrase \textit{and no one/nothing else}. This is the example from section 1.4:

\[(48)\]  
\[Q \text{ Who stole all the pens from your classroom?} \]
\[A \text{ John did.} \]

The answer in (48) is adequately paraphrased as "it was John, and no one else, who stole all the pens from my classroom". Now the point is that affirmative and negative answers to a topicalized yes/no-question can be adequately paraphrased in this way, but not the affirmative
and negative answers to a neutral yes/no-question. Consider the following examples involving affirmative answers to neutral yes/no-questions:

(49) Q Did you clean up your room?  
     A1 Yes (I did).

(50) Q Would you like a cup of coffee?  
     A1 Yes please.

The affirmative answers to the questions in (49) and (50) are not adequately paraphrased as "what I did was clean up my room, and nothing else" and "what I would like is a cup of coffee, and nothing else", respectively. An affirmative answer to these questions does not warrant any further conclusions about what the answerer did other than clean up his room or what he would like other than a cup of coffee. The same observation (exhaustive paraphrases are not possible) holds for negative answers to these neutral yes/no-questions:

(49)' Q Did you clean up your room?  
     A2 No (I didn't).
(50)' Q Would you like a cup of coffee?  
     A2 No thank you.

These negative answers are not adequately paraphrased as "what I did is not clean up my room; I did all kinds of other things" and "what I want is not a cup of coffee, but I would like something else", respectively. Again, any conclusions which go further than "I did not clean up my room" and "I would not like a cup of coffee" are not warranted on the basis of the negative answer to these yes/no-questions.

On the other hand, answers to topicalized yes/no-questions do allow for an exhaustive interpretation. Some examples are

(51) Q Did you finish your BOOK in time?  
     A1 Yes I did.  
     A2 No I didn't.
(52) Q Did you find your SHOES?  
     A1 Yes I did.  
     A2 No I didn't.

Here, the affirmative answers may be paraphrased as "what I did finish in time was my book, but not anything else" and "what I found were my shoes, but I didn't find anything else". The negative answers may be paraphrased as "what I finished in time was not my book, but something else", and "what I found were not my shoes, but something else". The role of exhaustive interpretation in topicalized yes/no-questions is better seen in cases where the focused element does not coincide with the element standardly receiving the main sentence accent, as in this example:

(53) Did you find YOUR shoes?

---

54 Indirect answers to yes/no-questions are not considered here, and neither are yes/no-questions with multiple focused elements (see Hirschberg 1991:146).
Both an affirmative and a negative answer can be paraphrased as exhaustive: "what I found was my own shoes, but no-one else's (e.g. not Mary's)" and "what I found was not my own shoes, but I did find someone else's (e.g. Mary's)", respectively. Thus, with respect to exhaustive interpretation of the answer, topicalized yes/no-questions are comparable to WH-questions.

The fact that the difference between neutral and topicalized yes/no-questions was not always recognized has given rise to problems with respect to the interpretation of numerals on more than one occasion. The first problem is signalled by Horn (1972:33). In (54) we have an example of a yes/no-question, to which two opposing answers are equally possible:

(54) Q Does John have four sheep?
    A1 Yes, (in fact) he has five.
    A2 No, he has five.

In (54), both A1 and A2 are adequate answers to the question. But what determines the choice for either of these? Horn notes that "contextual clues" are responsible, but does not identify these contextual clues. With the analysis presented here that is possible: the question in (54) is the surface form of at least two different questions, namely a neutral yes/no-question in (55), and a topicalized yes/no-question in (56). Answer A1 is only acceptable in the analysis under (55), which represents the existential reading of the numeral:

(55) Q Does John have four sheep?
    Are there four sheep that John owns?
    A Yes, (in fact) he has five.

Answer A2 is only acceptable under a different analysis, incorporating the cardinality reading of the numeral in a topicalized yes/no-question:

(56) Q Does John have four sheep?
    Is the number of sheep that John owns four?
    A No, he has five.
    The number of sheep that John owns is five.

Because the numeral receives the cardinality reading, a correction is necessary: No, he has five.

A similar problem arises when yes/no-questions are used as a test for lexical ambiguity. In a chapter on lexical ambiguity, Cruse (1986:64/5) includes numerals in a section named "some difficult cases". Numerals would on some tests come out as ambiguous between an "at least" and an "exactly" reading, but not on others. Specifically, on what Cruse calls the yes/no-test, numerals would be lexically ambiguous. Cruse's example is similar to Horn's example:

(57) Q Have you got £10 in your wallet?
    A1 Yes. In fact, I've got £12.
    A2 No, I've got £12.
According to Cruse, these two different answers would result from the fact that the respondent reacts to the sense he believes the questioner to be intending. The criterion tested for is that "separate senses should be independently maximisable" (1986:60), but it is not very clear what Cruse means by this. However, it is clear that a different analysis of this example is available: the two answers result from the fact that the information contained in the question can be structured in several different ways, among which are the following two structures (corresponding to the existential and the cardinality reading of the numeral, respectively):

(58) Q Have you got £10 in your wallet?
   Is there £10 in your wallet? [neutral yes/no-question]
   A Yes. In fact, I've got £12.

(59) Q Have you got £10 in your wallet?
   [the amount of money in your wallet] be[^sp] [£10]?
   Is the amount of money in your wallet £10? [topicalized yes/no-question]
   A No, I've got £12.

(58) gives a neutral yes/no-question, which gives the existential reading of the numeral, while in (59) we have a topicalized yes/no-question with the cardinality reading of the numeral. On this analysis, there is no lexical ambiguity of the numerals, but an ambiguity in the information structure of the question. Consequently, the yes/no-test is, certainly without any further specification, not suitable as a test for lexical ambiguity. The source for this is the ambiguity of information structure incorporated in the yes/no-question. Therefore, a 'positive' result on this test (i.e., both an affirmative and a negative answer are possible to the same question) has two possible causes: either a lexical ambiguity or an ambiguity of the information structure in the question. The test by itself is not capable of distinguishing between these two.

3.6 Conclusion

The arguments presented in this chapter allow for several conclusions, both general and specific. To start with the general conclusions, a sentence can be used to express several utterances (tokens) which differ in their information structure. Regarding sentences as answers to questions, thus turning them into utterances, provides an adequate method of bringing out these differences in information structure. Neither of these observations is new; both were made by the philosopher Stout one hundred years ago (1896:214):

All answers to questions are, as such, predicates, and all predicates may be regarded as answers to possible questions. If the statement, "I am hungry" be a reply to the question, "Who is hungry?", then "I" is the predicate. If it be the answer to the question, "Is there anything amiss with you?" then "hungry" is the predicate. If the question is, "Are you really hungry?" then "am" is the predicate. Every fresh step in a train of thought may be regarded as an answer to a question. The subject is, so to speak, the formulation of the question; the predicate is the answer.
A conclusion which relates to present-day discussions lies in the fact that these differences in information structure can have truth-conditional consequences, as was shown here with respect to the interpretation of numerals. This brings us to the specific conclusions of this chapter.

We have observed that numerals allow for three different readings, which are often indicated as the "at least", the "at most" and the "exactly" readings. The cases where numerals indicate an upper limit, or receive the "at most" reading, were all seen to involve a numeral in the scope of a modal predicate (e.g. can or an operator indicating an upper limit, e.g. enough). The cases where numerals indicate a lower limit, or receive the "at least" reading, fall into two different categories: one where the numeral is in the scope of a modal or limiting predicate (e.g. criteria and requirements), and one in which the numeral, due to its position in the information structure of the utterance, receives the existential reading. The cases involving modal or limiting predicates are considered to lie outside the scope of this study.

The dominant view on the semantics and pragmatics of numerals is that their truth-conditional meaning is "at least", which is accompanied by a generalized implicature "at most". These together lead to the desired interpretation. Apart from the theoretical objections against implicature, which were discussed in the previous chapter, there are also descriptive problems for this view. It leads e.g. to the undesirable consequence that at least n and n are synonymous. When the effects of information structure are taken into account, it is possible to give an adequate unitary semantics for numerals. No appeal to theoretically unstable pragmatic inferences is necessary. Semantically, numerals are predicates indicating a cardinality, a definite quantity. To put it somewhat crudely, two means "two" under all circumstances. It is the information structure of an utterance which is responsible for the fact that numerals appear as if they receive two different readings. When a numeral is in comment position, i.e. when it functions as the answer to a WH-question, it receives what we have called the exhaustive or cardinality reading ("exactly"). When the first occurrence of a numeral is in non-comment position it receives the existential reading ("at least"). The latter reading is analyzed as involving the existential quantifier as the highest predicate in the SA. The former reading has, at the level of the SA, the numeral in the NP\textsubscript{2} position with the abstract predicate bes\textsubscript{p} as highest predicate in the SA. In chapter 4 I will, after discussing another group of classical generalized Q1 implicature examples and analyzing them in the framework adopted here, discuss the exhaustive reading in more detail.
4 Exhaustive Interpretation in Discourse

4.1 Introduction

Having proposed a new account of the semantics and pragmatics of numerals in the previous chapter, we shall now look at other cases of the exclusion inference in a different light. In this chapter I will argue that the exclusion inference is in fact built from two separate processes: a process of exhaustive interpretation of all lexical material in comment position, which is the main topic of the present chapter, and a process of "negating the complement", which I will discuss and argue for briefly in section 4.3.4.

In arguing these claims, I will be challenging two assumptions about Q1 implicatures, as understood by Radical Pragmatics and its followers. They were pointed out in chapter 2: the Gricean argument formulated there reveals two central assumptions of the theory of scalar implicature as developed by Radical Pragmatics, namely its assumptions with respect to the semantics of scalar terms, and the fact that it is the availability of some expression alternative that triggers the implicature. I hope to show that the facts covered by the theory of generalized Q1 implicatures are better explained differently, dropping both these assumptions. First, there is a semantic process called exhaustive interpretation (attributing a value), which invariably operates on lexical material in comment-position, independently therefore of the presence of any sensible alternative. This interpretation is incorporated into the semantic structure of the utterance, as represented by the SA. It is thus not a separate semantic or pragmatic operation, performed after establishing the meaning (in a strict sense) of an utterance. Secondly, the semantics of the scalar terms does not include the "at least" condition that Radical Pragmatics is forced to assume.

The central issue in this chapter is the interaction between the exclusion inference that an utterance induces, and the context in which it appears, i.e. the question to which it is supposed to be an answer. The relation between (explicit) questions and exclusion inferences is in fact considered in implicature theory, but in an entirely different way. The relationship between (explicit) questions and answers is considered, by Radical Pragmatics, to induce exclusion inferences in a particular way. Questions express a request for a certain specified amount of information (or alternatively some action on the part of the hearer), and if the answer does not comply with this request (if it is not satisfactory), then this is meaningful: it induces an inference with the general form “the speaker does not know the answer to the question” or “the speaker is not able/willing to provide the information asked for”. E.g., there is Grice’s famous (though somewhat defective) example

A is planning with B an itinerary for a holiday in France. Both know that A wants to see his friend C, it to do so would not involve too great a prolongation of his journey:

1. A Where does C live?
   B Somewhere in the south of France.

(Grice 1989:32)

The information asked for in this context is the name of the town in which C lives, so that the trip can be planned adequately. B’s answer does not provide this information (is not
informative enough, thus violates maxim Q1) and thus leads to the inference “B does not know in which town C lives”.\textsuperscript{55}

In this chapter the focus will be on the influence of the information structure of an utterance (as determined by the question to which it is supposed to be an answer) on the inferences induced by that utterance. It will be shown, just as in the previous chapter, that there is one strong claim of Radical Pragmatics that does not hold: the exclusion inference, which by definition should be cancellable (i.e., negating the inference does not result in a contradiction), very often is not. When the concepts inducing the inference are in comment position, negating the inference leads to a contradiction. This serves as evidence that the process of interpretation at work here is not Q1 implicature as understood by Radical Pragmatics. The alternative proposed here, exhaustive interpretation, is discussed in detail in section 4.3.

Section 4.2 contains a detailed survey of a large number of examples, many of which are taken from the Radical Pragmatics literature. The aim is to show that, when the lexical item inducing the exclusion inference is in comment position, cancellation of such an exclusion inference leads to an unacceptable piece of discourse. In section 4.3 the alternative explanation of the exclusion inference will be discussed in detail; the theoretic context as sketched in the earlier sections of chapter 3 plays an important role here. In section 4.4 this proposal will be discussed in comparison with another theory of exhaustive interpretation. Finally, in section 4.5, I will show that the theory developed here provides a way out of a dilemma posed by Harnish (1976) and a similar dilemma posed by Horn.

4.2 Information structure and the exclusion inference

Subsections 4.2.1 and 4.2.2, which discuss exclusion inferences induced by lexical material in comment position and in non-comment position, respectively, are structured identically, in order to facilitate a comparison. In each of these sections, subsections are devoted to separate categories of lexical items claimed to induce a Q1 implicature: mutually exclusive items and scalar predicates. The latter category involves notably the scalar adjectives. The former category, that of the mutually exclusive terms, is a special case. Often in the literature, examples involving a mutually exclusive term inducing a Q1 implicature are given, but no author makes the general claim that the whole category of mutually exclusive terms induces Q1 implicatures (with the exception of Hirschberg, see the discussion in section 2.1.2).

4.2.1 Lexical material in comment position

In virtually all examples of utterances inducing an exclusion inference in the Radical Pragmatics literature, the most likely reading is such that the lexical item that triggers the exclusion inference is in comment position. This is significant because, in such a constellation, applying the cancellation test, which as we saw in chapter 2 is virtually the only test for establishing the presence of an exclusion inference, very often does not lead to an acceptable piece of discourse. In this section I will review a number of examples from the literature, establishing first their information structure in the context of a question and

\textsuperscript{55} The problem with this example is that it is not very convincing. If A and C are friends, how can it be that A doesn't know in which town C lives?
secondly the acceptability of examples of cancellation of the exclusion inference in the given context. The semantics of the so-called cancellation phrases, which is generally neglected in the Radical Pragmatics literature, plays an important role in this.

4.2.1.1 Mutually exclusive predicates
In chapter 3 we saw that cancellation of exclusion inferences induced by numerals leads to straightforward inacceptability when the numeral functions as an answer to a *how many* question, thus when the numeral is in comment position:

(2) <How many children does John have?>
   a. He has fourteen. *In fact he has twenty.*
   b. Not just fourteen, but fifteen.
   c. Not fourteen, but twelve.

Now (2b) and (c) are examples of alleged cancellation which are only acceptable under very specific circumstances: when it was stated or suggested earlier on in the discourse that John (possibly) has (exactly) 14 children. In this example, and in many that will follow in this section, the semantics of the phrase *not (just) P but Q* plays an important role. Due to insufficient attention to the semantics of this phrase, it was never recognized that it introduces what one could call a *quote answer*. It is only acceptable when *P* is a quote, that is when it was uttered or suggested earlier on in the discourse. One could say that the occurrence of *P* is a presupposition of the phrase *not (just) P but Q*. *Q* then serves as a correction on *P*. The character of *not (just) P but Q* as a quote answer makes it unjustified to employ it as an argument for the existence of a scale as an abstract semantic entity (see the discussion in section 2.1.6).

In discussing the above example I have taken the liberty of putting *just* in brackets. This is not what Radical Pragmatics does, but there are good reasons for including *not P but Q* in a discussion of cancellation frames, and including examples like (2c). It is obvious why RP would want to exclude it: RP predicts an asymmetry between, on the one hand, replacing a cardinal number by a higher number (as in (2b)), and on the other hand, replacing a cardinal number by a lower number (as in (2c)). The former is considered to be a case of implicature cancellation, the latter presumably a case of correction. However, given the semantics of *not (just) P but Q* it is better to view both of them as cases of correction, where what is being corrected is required to be present in the context. The difference between (2b) and (2c) then is that the use of *just* in (2b) signals that the correction, the new answer, is more informative (in the sense defined in chapter 3) than the old answer.

Two other examples from Levinson (1983:106), not involving numerals, show the same kind of unacceptability as (2a):

---

56 This example is adapted from Levinson (1983:106).
(3) <What colour is the flag?>
   <What does the flag look like?>
   a. The flag is white.
   b. The flag is white. *In fact it is red.
   c. The flag is not only white, but red as well.
   d. The flag is white, with red stars on it.

(3a), which is Levinson's original example (without the question!), is taken to implicate that the flag has no other colours and is thus entirely white. Given this interpretation, the implicit question can only be as given here. On any question placing white in the non-comment expression this inference is not present:

(4) Mother A and child B (age 4) are playing a game. The child is asked to identify white objects.
   A       What more is white?
   B       The flag is white.

In this context, uttering the flag is white does not exclude the possibility that it has e.g. dots or stripes of some other colour. At most it can be taken to mean that the flag is largely or predominantly white. Note moreover that B in (4) cannot be taken to imply that the flag is the only thing that is white; the flag is not intended as an exhaustive listing of everything that is white. This is due to the context, reflected in the question.

Example (3b), which employs a different cancellation frame, is not an adequate answer to either of the questions given, and thus does not result in an acceptable piece of discourse. This inacceptability results from the semantics of in fact: the use of in fact signals that what follows is a further specification of what was said before. It signals a further restriction of the set of possible situations that will satisfy the utterance; thus, it signals that the information following is more informative than what was said before. In general, if someone utters P, in fact Q, then Q is required to be more informative than P.57 The answer in (3b) does not meet this requirement: the flag is red is on no account more informative or specific than the flag is white. Use of the cancellation frame P, in fact Q will never result in an acceptable piece of discourse when the alternates P and Q are mutually exclusive. Moreover, and this argument will reappear in other cases, where the requirements are in fact met, from a discourse point of view it is odd for a speaker to give a particular answer or make a particular statement (The flag is white) and immediately replace it with a different statement (The flag is red). If it is intended as a case of correction, it should be adequately marked as such.

The cancellation in (3c) is acceptable, but only when it was suggested or claimed previously that the flag is white. The phrase not only P but Q (as well) is again a quote answer, which presupposes the presence of P in the wider context. Again, no abstract notion of scale needs to be invoked, because in the context of question (3) answer (3c) would not be an adequate contribution to a discourse in which the possibility or the fact that the flag is white had not yet been introduced. In order to account for (3c) it is thus not necessary to hypothesize a scale of colour terms, as Levinson suggests. Finally, (3d) would in Radical Pragmatics presumably be viewed as a case of cancellation as well, where the phrase with red stars on it cancels the inference "only or entirely white". However, (3d) is better viewed in its entirety as an answer to the question in (3). To resume, the exclusion inference in (3a) is not a

57 This is consistent with what Radical Pragmatics claims with respect to the sentence frame P, in fact Q. See e.g. Horn (1972:40).
result of the use of the word *white* in the presence of available alternatives such as *blue, black* etc., but is due to the place that predicate obtains in the information structure of the utterance, which is essentially due to the question. Moreover, the cancellation cases in (3b-d) are not indicative of a scale as some abstract semantic entity.

A further example from Levinson (1983:106) is

(5) How did Harry fare in court the other day?
   a. Oh he got a fine.
   b. Oh *he got a fine. *In fact he got a prison sentence.
   c. Oh he got a fine. No wait, he got a prison sentence.
   d. He got a fine, and a few days in prison.

In (5), the question and answer (a) are the original example. Levinson remarks that

> If it later transpires that Harry got a life sentence too, then B (if he knew this all along) would certainly be guilty of misleading A, for he has failed to provide all the information that might reasonably be required in the situation. (1983:106)

(5b) contains a straightforward cancellation of the exclusion inference that Harry got only a fine, a fine and nothing else; it is unacceptable as an answer to the question posed, for the same two reasons as (3b) was unacceptable: it can only be coherently interpreted as a case of correction, but it is not marked as such, and secondly the utterance is in conflict with the semantics of *in fact*.58 In order to make this an acceptable case of correction, it needs to be marked as such, as in (5c). Finally, (5d) is similar to (5a), but states that Harry was punished not in one way, as in (5a), but in two different ways. (5d) induces the inference, brought forth by the interaction between question and answer, that Harry got nothing more than a fine and a few days in prison. That is, the hearer will interpret both (5a) and (5d) as exhaustive answers to the question in (5).

A different example exhibiting similar phenomena comes from Carston (1985:3):

(6) *<Who will she listen to?>*
   a. She’ll listen to people who propound left-wing views.
   b. She won’t just listen to people who propound left-wing views, she’ll listen to anybody who has something interesting to say.
   c. She’ll listen to people who propound left-wing views, and to anyone else who has something interesting to say.

(6a), which is (in isolation) Carston's original example, is said to induce the Q1 implicature "She won’t listen to people who put forward anything other than left-wing views". This example is similar to example (3) above, in that this implicature will only appear in the given

58 The examples (3) and (5), though both from the Radical Pragmatics literature, are generally considered to be marginal cases of implicature, because they involve logically independent terms as triggers for the Q1 implicature (exclusion inference). However, if one sticks to the cancellation frames as indicative of the presence of an implicature, then the argument presented here with respect to *in fact* provides an argument for discarding mutually exclusive terms from the group of implicature-triggering concepts. It is clear that I do not advocate such a move, but on the contrary wish to include these mutually exclusive concepts into my study, because I think that they show most clearly the process of exhaustive interpretation, which is a structural property of the interpretation of an utterance, and is not related to particular concepts.
context. If the implicature-inducing expression people who propound left-wing views appears in non-comment position, as in (7), the implicature is not induced:

(7)  
A  Who will listen to people who propound left-wing views?  
B  Mary will.

In (7), the possibility that Mary will listen to people who express different views is not excluded. Again, it is the interaction between question and answer that is responsible for the inference in (6a). The cancellation examples (b-c) are similar to previous examples: (6b) is a quote answer, indicating an ordering by informativity. In (6b), the set of possible situations satisfying she’ll listen to anybody who has something interesting to say is considered by the speaker to contain a proper subset the set of possible situations satisfying she’ll listen to people who propound left-wing views. (6c) is similar to (3d) and (5d) and constitutes in its entirety an exhaustive answer to the question. Again in this example, as in the previous ones, it is the interaction with the question that is responsible for the fact that the expression people who propound left-wing views leads to the implicit message (She won’t listen to) people who do not propound left-wing views.

Finally two examples from Kempson (1986:81/2), who includes these cases in "the whole set of scalar implicatures":

(8) She didn't lose a finger, she lost an arm.
(9) I didn't invite John to supper: I invited John, Mary and Susanna.

Both are alleged examples of implicature cancellation, and in both cases the claims made with respect to the previous examples hold. These utterances are only acceptable when the alternatives being denied, namely finger or lose a finger in (8) and John in (9), were previously mentioned as answer to a question (i.e. functioning as comment expression):

(8)' <What did she lose?>
A  She lost a finger.
B  She didn't lose a finger, she lost an arm.

(8)' <What happened to her in the accident?>
A  She lost a finger.
B  She didn't lose a finger, she lost an arm.

(9)' <Who did you invite to supper? John?>
I didn't invite John to supper: I invited John, Mary and Susanna.

Examples (3-9) were all somewhat atypical cases of implicature, as the expression alternatives (e.g. in (3) white vs. red, blue, black etc.; in (5) getting a fine vs. going to prison, in (6) people who propound left-wing views vs people who do not propound left-wing views) do not constitute a scale in the classical sense (they are not ordered by one-sided entailment). They do fit into Hirschberg's more liberal description of orderings supporting Q1 implicature.

4.2.1.2 Scalar predicates
The more familiar examples of generalized Q1 (scalar) implicatures can be looked at as well in the perspective developed above. The following examples all have the same structure: first the scale is mentioned that is claimed to be responsible for the generation of a particular
exclusion inference, next the question that represents the context (which is formed in such a way that the lexical item (concept) inducing the exclusion inference is in comment position). The a-sentence is the utterance inducing the exclusion inference, the b-sentence is the explicit version of the exclusion inference, the c- and d-sentences are cases of cancellation.

(10) <sometimes, often, always/every day>
    How often do you visit your grandmother?
    a. I visit her often.
    b. I do not visit her every day.
    c. I visit her often, and in fact every day.
    d. I don't just visit her often, I visit her every day.

(11) <possible, probable, certain>
    How likely is it that we will finish in time?
    a. It's probable.
    b. It's not certain.
    c. It's probable, and in fact certain.
    d. It's not just probable/likely, it's certain.

(12) <some, many, all>
    How many pupils failed the test?
    a. Some.
    b1. Not all pupils failed the test.
    b2. The speaker doesn't know the exact number of pupils that failed the test.
    c. Some, in fact all.
    d. Not just some, but all.

In (12), both (b1) and (b2) are derived according to Grice's Q1 maxim; (b1) is the scalar implicature, dependent on the scale mentioned, (b2) has the general format of an implicature induced by an utterance that was not "informative enough for the purposes of the exchange". In the discussion to follow, I will concentrate on (b1).

(13) <cool, cold, freezing>
    How's the water?
    Why can't I go for a swim?
    a. The water is cold.
    b. The water isn't freezing.
    c. The water is cold, in fact it's freezing.
    d. The water is not just cold, but freezing.

(14) <pretty, beautiful, gorgeous>
    What do you think of him?
    a. He's beautiful.
    b. He's not gorgeous.
    c. He's beautiful, in fact gorgeous.
    d. He's not just beautiful, he's gorgeous.
The cancellation-examples in the c-sentences are all formed with the sentence frame *in fact*. In these cases, there is no conflict with the semantics of *in fact* which could give rise to an unacceptable piece of discourse, as was the case in the previous subsection. On the contrary, when the utterances containing an *in fact*-phrase are considered in their entirety to be answers to the question given, there is no unacceptability. There is however a phenomenon related to discourse segmentation. The c-sentences can be regarded as giving two different answers to the same question: *How often do you visit your grandmother?* *I visit her often/I visit her every day.* This is likely to raise eyebrows: why would someone give an answer to a question and immediately replace it with another, more informative one? This is what makes the interpretation a rhetorical figure, drawing the attention to the extreme qualities of the entity under discussion (e.g., stressing the frequency with which the speaker visits his or her grandmother).

The d-sentences are all quote answers, only acceptable if the first answer is present in previous discourse. The fact that the second answer is more informative than the first is required by the semantics of *not just P, but Q*. However, *just* can be left out without altering the acceptability of the answers:

(10) d'. *I don't visit her often, I visit her every day.*
(11) d'. *It's not probable/likely, it's certain.*
(12) d'. *Not some, but all.*
(13) d'. *The water is not cold, but freezing.*
(14) d'. *He's not beautiful, he's gorgeous.*

With this alteration, the fact that we are dealing here with a correction on a previously given characterisation, and not with implicature cancellation, is more clearly exhibited.

Obviously, the examples concerning mutually exclusive items, discussed in the previous subsection, show a different pattern from the examples concerning classical cases of scalar terms. In this latter category, employing the sentence frames *in fact* and *not just/only* in order to correct a qualification given in comment position does not lead to an unacceptable piece of discourse. This observation may be interpreted in various ways. First of all, it may be taken as argument for the fact that the scalar predicates form a special category, semantically speaking. However, this conclusion should not be drawn too hastily. Even the colour terms, which in the discussion above have figured as a primary example of the category of mutually exclusive terms, may enter into a context where there may be discussion over the various nuances of a colour; in such a context, they side with the scalar predicates, not with the mutually exclusive ones:

(15) *What colour is that curtain?*
*It is red, in fact it is magenta.*

In other words, the categorization made above, i.e. the distinction between mutually exclusive predicates and scalar predicates, may be intuitively clear but at least has some exceptions. The second conclusion that can be drawn from the observations in this section is that the examples involving scalar terms are only acceptable when regarded as answers in their entirety. Thus interpreted, they do have a rhetorical flavor to them.

Finally, another observation is in order here as well. Recall the structure of implicature theory: it is the availability of an alternative, stronger expression which, when a weaker but compatible expression is used, leads to the implicit denial of that stronger expression. This
implicit denial has the status of conversational implicature, that is, a non-truth-conditional inference. The fact that this inference is non-truth-conditional is supported by giving examples like the c- and d-sentences above, which are regarded as examples of cancellation of the implicit inference. Within the theoretic context of implicature theory it is indeed possible to view these examples as such, but apart from that theoretic context I believe that it is an unnecessarily complicated way of regarding the examples given. The sentence frames \( P, \text{in fact } Q \) and \( \text{not just/only } P, \text{but } Q \) (as well) each have their own specific semantics. \( P, \text{in fact } Q \) demands that the information given by \( Q \) is more informative than the information given by \( P \). This accounts for the fact that alternate expressions \( P \) and \( Q \) between which there is no ordering of relative informativeness result in an unacceptable piece of discourse (as was shown in section 4.2.1.1). No theoretic assumptions with respect to implicature cancellation are necessary to account for this. Secondly, the sentence frame \( \text{not just/only } P, \text{but } Q \) (as well) is a close relative of \( \text{not } P, \text{but } Q \). This is a quote answer, which has as a presupposition the presence of \( P \) in the previous context. It is the presence of \( \text{just or only } \) which makes for the requirement that \( Q \) be more informative than \( P \); without \( \text{just or only } \), this sentence frame simply represents a correction. Again, no assumptions with respect to implicature cancellation are necessary to account for this sentence frame. Consequently, in the discussion to follow the sentence frames \( P, \text{in fact } Q \) and \( \text{not just/only } P \) will be referred to as cases of correction.

4.2.2 Lexical material in non-comment position

4.2.2.1 Mutually exclusive predicates
The examples to be discussed in this section are all variants on the examples introduced in the previous section. With respect to the interpretation of the colour terms we saw in example (4), repeated here, that in the context as given below, \( \text{white} \) does not exclude the possibility that the flag contains other colours:

(16) Mother and child (age 4) are playing a game. The child is asked to identify white objects.
    mother What more is white?
    child The flag is white.

Cancellation of the alleged exclusion inference, \( \text{the flag is wholly white} \), is unproblematic, contrary to what we saw in section 4.2.1 (example (3)):

(17) mother What more is white?
    child The flag is white. There's some blue in it, too.

One may object to this argument that the child, by adding this phrase, has shifted the topic of the conversation. The original question defines the set of things that are white as topic, but topic is now shifted to the colours that the flag contains. A more adequate representation of the discourse in (17) is (17)′:

(17)' mother What more is white?
    child The flag is white.
    <What more colours does the flag have?>
    There's some blue in it, too.
Nevertheless, this topic shift does not affect the point made here: cancellation of exclusion inferences induced by concepts (originally) in topic position is unproblematic. This will invariably constitute a topic shift, sidestepping the original issue, but it does not lead to incoherent or unacceptable discourse. The topic shift involves an additional implicit question, as will become clear in the examples to follow. In (18) we see the same constellation as in (17):

(18) A Who has five guilders?/ Who can lend me five guilders?
    B John has/can.
    A How much can he lend me?
    B He even has/can lend you six.

(19) A group of friends was arrested after a hot night in town. The next morning they are arraigned in court. Two people are discussing the sentences given.
    A Who got a fine?
    B Harry got a fine.
    A What more did he get?
    B He got a short prison sentence (as well).

In (20), the correction is no problem, but on the same condition that we saw in the previous section: B in effect claims that the set of people with strong political views includes the set of people who propound left wing views as a strict subset, due to the semantics of in fact:

(20) A Who will listen to people who propound left-wing views?
    B Mary will.
    <A To whom will she listen?><
    B In fact she'll listen to anybody who has strong political views.

The analysis given here is thus as follows: alleged cases of cancellation of an exclusion inference induced by lexical items in non-comment position in fact are cases of correction, involving a topic shift. The topic shift is signalled by the second question.

4.2.2.2 Scalar predicates

In this subsection I will present a number of examples concerning scalar predicates in non-comment position. The aim is to show that, just as in the examples with mutually exclusive terms, correction is not problematic at all, but standardly involves a topic shift. The examples in this subsection are mirror images of those in subsection 4.2.1.2. I will first mention the relevant scale again, and then present a small piece of discourse involving correction.

(21) <sometimes, often, always/every day>
    A Who comes to visit you often?
    B My daughter.
    <A How often does she visit you?><
    B She visits me every day.

(22) <possible, probable, certain>
    A Who thinks that it is possible to finish in time?
B I do.

<A How likely do you think it is that we will finish in time?>
B In fact I'm sure that we will finish in time.

(23) <some, many, all>
A Why did some of your pupils fail?
B Their work was a complete mess.
<A How many failed?>
B In fact they all failed.

In the above examples, just like the examples involving mutually exclusive cases, the corrected information (every day, I'm sure, all) is again in comment position and is therefore interpreted exhaustively. Any attempt to alter or correct this information would require adequate marking as such. The same holds for the following examples:

(24) <pretty, beautiful, gorgeous>
A Why do you find her beautiful?
B I just love her eyes.
<A How would you describe her?>
B She's gorgeous.

(25) <want to X, try to X, succeed in X-ing>
A What did you try to do?
B I tried to get the book out of the pool.
<A Did you succeed?>
B And I succeeded, too.

(26) <like, love, adore>
A Do you like your teacher?
B I do.
<A What are your feelings towards him?>
B I love him.

The results of the survey conducted in this section may be summarized as follows. First of all, the phrases P, in fact Q and not (just/only) P, but Q (as well) are more adequately characterized as corrections instead of cancellations. Secondly, the possibility of corrections is related to the place that a piece of information occupies in the information structure of the discourse. When the relevant piece of information is a mutually exclusive concept in comment position, correction by means of in fact is hard or even impossible. For the scalar predicates in comment position this form of correction is possible, but only when the whole phrase P, in fact Q functions as a single answer. This answer then does have a rhetorical flavour to it. Finally, when the piece of information to be corrected is in non-comment position correction is always possible, but this necessarily involves a topic shift.

The explanation for the fact that correction in comment position is sometimes impossible, sometimes merely hard, but never straightforwardly easy, lies in the fact that any lexical material in comment position is interpreted exhaustively. Before discussing the details of this process more thoroughly (supplementary to the discussion in chapter 3) I will first discuss another context in which the exclusion inference systematically occurs.
4.2.3 Restrictive and nonrestrictive relative clauses

In the context of the investigation conducted here special attention must be paid to the difference between restrictive and nonrestrictive relative clauses. The inferences brought about by restrictive clauses are sometimes regarded as an instance of implicature (notably by Carston (1985); cf. example (6) above), but there are some problems regarding this point. Restrictive relative clauses induce exclusion inferences, but especially their cancellability is doubtful, which is problematic for Radical Pragmatics. On the views developed in this book, there is no problem at all.

When restrictive and nonrestrictive relative clauses are looked at in the perspective as adopted in this book, they are seen to possess two remarkable features. First of all, supplying a question as context exhibits their information structure and thus identifies the difference between restrictive and nonrestrictive relative clauses. Secondly, only restrictive relative clauses characteristically induce an exclusion inference.

(27) The men who were on their way home got lost in the forest.
(28) The men, who were on their way home, got lost in the forest.

(27) contains a restrictive, (28) a nonrestrictive clause. Only (27) invites the characteristic inference "the men who weren't on their way home did not get lost in the forest". (28) does not give rise to such an inference at all.

The question arises why this is so. This can be accounted for by looking at these sentences in context, more specifically in the context of questions to which they would form an appropriate answer. Looked at this way, it turns out that (27) and (28) are structured differently:

(27)' <Who got lost in the forest?>
    The men who were on their way home (got lost in the forest).

(28)' <What happened?>
    The men,
        <What were they doing?>
        (who) were on their way home,
        got lost in the forest.

Thus analyzed, the restrictive clause in (27) is part of the answer to a topic-forming question, the nonrestrictive clause in (28) is the answer to a subquestion. The discourse in (28)' contains two question-answer pairs, which can also be ordered differently without changing the content of the discourse:

(28)"<What happened?>
    The men got lost in the forest.
    <What were they doing?>
They were on their way home. 59

Problems with cancellation (the alleged implicature cannot really be cancelled) do not arise in this analysis: the expression giving rise to the inference is part of the comment-expression, and this makes the inference truth-conditional and therefore not cancellable. (27) can, keeping the truth-conditions identical, be paraphrased as Only the men who were on their way home, and nobody else who is relevant in the situation at hand, got lost in the forest.

4.3 Exhaustive Interpretation

4.3.1 Two properties of exhaustive interpretation

We have established, in the previous sections, that the exclusion inference interacts with discourse structure in such a way that any exclusion inference induced by lexical items in comment position is truth-conditional. In chapter 3 we suggested as alternative explanation the process of exhaustive interpretation, which we will develop more fully now. In this section I will first present some considerations establishing two properties of exhaustive interpretation, i.e. it is truth-conditional and it is a relational notion, and then show how these properties are integrated into the semantics of questions and answers as presented in chapter 3.

Informally speaking, exhaustiveness is a property of the interpretation of utterances and its workings can be described as follows: some aspects of an utterance indicate that it has to be interpreted exhaustively. When the hearer perceives these clues correctly, he will then draw the conclusion that the utterance covers the relevant (set of) objects/situations in the world completely. Thus, in (29),

(29) \[ A \quad \text{Who ruined your vegetable garden?} \\
B \quad \text{Our neighbour's dogs did.} \]

the speaker expresses, and the hearer concludes, that only the neighbour's dogs, and no other person or animal, ruined his vegetable garden. As said in section 1.4, an utterance which is interpreted exhaustively is interpreted as "speaking the whole truth", not as "speaking the truth" simpliciter. In other words, the fact that by an utterance some possibilities are discarded is truth-conditional: not only is some situation affirmed, but other possible situations are

59 This difference between restrictive and nonrestrictive subclauses is also observed in Van Kuppevelt (1991:53-55). He notes that the independent status of nonrestrictive subclauses in spoken discourse is signalled by intonation, as follows:

(i) <Who was arrested yesterday?>
Your friend JOHN was arrested yesterday. (restrictive)

(ii) <Who was arrested yesterday?>
Your FRIEND, <What's his name?>, JOHN, was arrested yesterday. (nonrestrictive)

60 In spoken discourse, intonation is the main indicator of the information structure (see Van Kuppevelt (1991) and the works discussed therein). Whether there are any decisive indicators in written discourse, apart from the reader's reconstruction of the information structure of the discourse (i.e., the reconstruction of implicit questions), is a matter that is open to further investigation.
(implicitly but truth-conditionally) negated. The following two examples come from a written biology test and illustrate the intuitive notion. Q is the test assignment, A is the answer given by a pupil (age 12):

(30) Q Mention a difference between bacteria and animal cells.
    A Bacteria do not have a nucleus.

(31) Q Mention a difference between bacteria and animals.
    A You can see bacteria only through a microscope.

What is conveyed here is, respectively, that animal cells do have a nucleus, and that animals can be seen without using a microscope. These answers are correct, although they are not completely explicit (or even stronger: not complete). The point is that there is, in the given context, no other implicit message possible: the test assignments restrict the relevant domain of discourse to a comparison between bacteria and animals or animal cells. Thus, when a property is attributed to bacteria, then, by exhaustive interpretation, the hearer (reader) will conclude that animals do not possess this property. The teacher will regard these as correct answers to the question, even though they are not complete.

A second characteristic of exhaustive interpretation is the fact that it is a relational notion; an utterance is only interpreted exhaustively in a context, i.e. as answer to a question. Utterance (32), uttered in isolation (e.g. functioning as a feeder (see section 1.3.2)), does not license any exhaustive interpretation:

(32) John hit Mary.

From (32), it cannot be concluded that only John hit Mary, or that only Mary was hit by John, or finally that all that John did to Mary was hit her. It says only that there was an event of John hitting Mary, and does not warrant any further conclusions. In (33) and (34), however,

(33) Q Who hit Mary?
    A John did.

(34) Q What did John do to Mary?
    A He hit her.

the topics are the un(der)specified set of persons that hit Mary, and the un(der)specified set of actions that John undertook with respect to Mary, respectively. In this context, the specifications asked for, i.e. the comment, is interpreted exhaustively: In (33), only John hit Mary (and no one else), and in (34), all that John did to Mary was hit her (and nothing else).

In order to capture both of these properties of exhaustive interpretation, i.e. its truth-conditional effects and the fact that it is a phenomenon related to questions and answers, no machinery is necessary additional to what was introduced in chapter 3. Both these properties of exhaustive interpretation are in fact integrated into the semantics of besp. The relation between questions and answers is reflected by the fact that questions and answers have the same SA. Any direct answer will fill in the empty spot in the SA; the WH-element functions as a variable. The fact that an exhaustive interpretation is truth-conditional follows from the characteristics of the abstract predicate besp: its satisfaction conditions demand that there be a function relating two variables. Technically, a function relates every argument to exactly one
element from the co-domain, every argument gets assigned exactly one value. Thus, when in discourse such a value is specified (by the answer to a question, the comment), its uniqueness is intrinsically asserted. Exhaustive interpretation is thus not a separate semantic or pragmatic operation, it is fully integrated into the semantics of questions and answers.

This view on exhaustive interpretation raises some problems, which I will address in the following subsections. For obviously, not every utterance is intended to be interpreted exhaustively. In the following subsections I will discuss various ways in which a speaker can prevent his utterance from being interpreted exhaustively, but first I will discuss two arguments against the semantic approach to exhaustivity presented here. First of all, there are questions which, at least at first sight, do not seem to ask for an exhaustive answer. An example can be found in this quote from Groenendijk and Stokhof (1984b):

(...) if you're walking down the road in your home-town and an Italian tourist addresses you, asking:

[35] Where can I buy an Italian newspaper?

you won't bore her citing a complete list of bookstalls and other places where Italian newspapers are sold. You just mention some place where she is likely to find one. And if you are a nice person you mention one that is not too far away and easy to find, and you won't try to be funny and answer *In Rome.*

The fact that (35) does not seem to ask for an exhaustive specification of places where one can buy an Italian newspaper can be explained as follows. It is a regular WH-question, but it seems reasonable to assume a pragmatic background principle stating that the answerer should name the nearest place where the speaker can buy an Italian newspaper. Evidently, the question (35) is asked because the speaker wants to buy an Italian newspaper during her stay in Amsterdam, and so an answer *In Rome* wouldn't be helpful, as wouldn't a complete listing of all places in Amsterdam where she could buy an Italian newspaper. The pragmatic background principle in fact becomes explicit sometimes, as it is not at all strange to answer as follows:

(35) Q Where can I buy an Italian newspaper?
A The nearest place is X, and there's also Y, but that's a little bit further away.

Thus this pragmatic principle can be seen to function, not so much as a correction on the exhaustivity, but as a principle limiting (in this particular case) the physical space within which exhaustivity is asked for.

A second argument comes from Beck and Rullmann (1995). They note that there are explicit exhaustivity markers in language, such as the German *alles* or the Dutch *allemaal*. They consider the existence of these markers as an argument against a semantic approach such as the one presented in this study, which makes all questions inherently exhaustive:

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: [36] differs in meaning from [37] in
that the former does not allow a nonexhaustive interpretation whereas the latter

does.

[36] Hans weiss wo man alles/überall die NYT kaufen kann.
Hans knows where one all/everywhere the NYT buy can
"Hans knows where all you can buy the NYT"

[37] Hans knows where you can buy the NYT.61

Although I agree with Beck and Rullmann that the existence of a genuine exhaustivity marker
would count as an argument against the approach developed here, I do not agree with them
that the Dutch allemaal (literally "everything, everyone") is such a marker (I limit myself to
Dutch, as this is my native language). The function of allemaal, both in questions and in
answers, rather seems to be to indicate that the speaker expects there to be more than one
object involved. A few examples may serve to illustrate this:

(38) a. Wat zit er allemaal in die doos?
What is all in that box?

b. Wat zit er in die doos?
What is in that box?

(39) a. Jan weet wat er allemaal in die doos zit.
Jan knows what all in that box is.
"Jan knows what's all in that box."

b. Jan weet wat er in die doos zit.
Jan knows what in that box is.
"Jan knows what's in that box."

Both in (38) and (39), the contrast between the a-sentences (with allemaal) and the b-
sentences (without allemaal) seems to be that the a-sentences express the suggestion or
expectation of the speaker that the contents of the box are not a single, large object, but rather
a multitude of small objects. This suggestion or expectation is absent from the b-sentences.
Moreover, in (39a), this suggestion is independent of the structural position of allemaal; it is
present both when (39a) is the answer to the question What does Jan know? and when it is the
answer to the question Who knows what's (all) in that box? Thus, the Dutch allemaal does not
function as an exhaustivity marker.

There is one very central aspect of the theory of implicatures which is conspicuously
missing from the alternative presented here: the generation of an implicature depends
crucially on the availability of an expression alternative. Determining which are, in a given
government, the expression alternatives of some utterance is also one of the main unsolved
problems for implicature theory. Exhaustive interpretation is an entirely different process; it is
a structural property of utterances in discourse, thus not limited to particular categories of
lexical items or semantic concepts (e.g., the scalar terms). However, this view of the process
of exhaustive interpretation that is responsible for the generation of exclusion inferences does
not account in any way for the selection of the relevant and meaningful expression

61 The paper from which this quote is taken does not include page numbers; this quote can be found in section
6.3, "(Non-)exhaustivity markers."
alternatives. I believe this to be a separate process, which I will discuss (though by no means account for fully) in section 4.3.3.

4.3.2 Escapes from exhaustive interpretation

In our notion of exhaustivity as described above, exhaustivity is inherent in any discourse. All utterances except feeders and context-neutral utterances are regarded as the answer to a question, be it implicit or explicit, and thus in any utterance there will be parts that are interpreted exhaustively. In explicit question-answer dialogues, the answerer is aware that (the appropriate parts of) his answers will inevitably be interpreted exhaustively. This is a very strong prediction, but it is not without escapes. There are various ways in which a speaker can prevent his utterance form being interpreted exhaustively: he can hedge his answer, indicating that it should not be regarded as exhaustive, he can object to the exhaustivity enclosed in the question (thus in a way refuse to answer the question) or he can correct an initial exhaustive answer. I will briefly discuss these strategies in turn; the characterizations will remain rather intuitive in nature. The aim of this discussion is to show that exhaustive interpretation, though integrated into the semantics of questions and answers, on the surface appears as a default strategy: any utterance will be interpreted exhaustively, unless the exhaustivity is explicitly cancelled.

4.3.2.1 Contradicting or avoiding exhaustivity

There are numerous modifications which a speaker can make to his utterance in order to prevent it from being interpreted exhaustively. The most explicit form is to contradict the exhaustiveness, i.e. to state that the answer is (possibly) not complete:

(40) Rather, I will simply list some whose comments, criticisms, collaboration, or support has been particularly salient – B.A., J.A., S.B., H.C., B. de C., R.C., ... – without implicating that the help provided by others was any more dispensable.  
(Horn 1989:xii; my emphasis, RS.)

However, a much more common way is to embed the answer under a higher operator: as far as I know, I think, I believe, thus avoiding exhaustivity:

(41) Q  What will happen to Boris in hospital?  
    A  As far as I know they will perform a bypass operation.

(42) Q  What did John do with his old car?  
    A  I believe he sold it to his neighbour.

In the context of a representational theory of semantics, these phrases have the function of opening an intensional subdomain in the truth domain. The truth domain is the main level of semantic representation; for any content represented here, the truth value must be assessed to test the truth value of what the speaker has said (the discourse). Any content represented in subdomains, though assessable for truth directly, does not have to be true for the discourse to be true. Thus, by embedding an answer which would otherwise be interpreted exhaustively the speaker avoids this commitment of giving a true and complete (exhaustive) answer. All
that can be assessed in (41) and (42) is the truth of the attribution of a certain knowledge or belief to the speaker.

The phrases exemplified in (41) and (42) are commonly regarded as invoking a particular type of Q1-implicature, namely a clausal implicature (Gazdar (1979)). Clausal implicatures, like the other categories of Q1 implicatures, depend on the contrast between semantically similar phrases, such as _I know_ and _I believe_. As stating _I know that P_ is stronger (more informative) than using _I believe that P_, the use of the latter is said to induce the Q1 implicature that the speaker is not sure of his answer: "the speaker doesn't know (for sure) that P". With the distinction between truth domains and subdomains in a representational semantics, the relevant contrast is different. The speaker can simply state something, whereby it becomes part of the truth domain and is up for assessment, or he can avoid the assessment of truth values by adding _I believe, I think_ etc., which open a subdomain. Thus, the fact that these phrases convey that the speaker is not sure of his answer results from the fact that they are embedded in a subdomain.

4.3.2.2 Objecting to the question: confusion regarding metalinguistic negation

The second way in which a speaker can prevent his utterance from being interpreted exhaustively is objecting to the question or the exhaustivity inherent in it. This strategy is probably most suited for explicit question-answer dialogues:

(43) Q1 Which T-shirt did you buy?
   A1 I didn't buy one, I bought two:
   <Q2 Which T-shirts did you buy?>
   A2 the red one and the yellow one.

In (43A1), the answerer does not answer the question but contradicts the assumption contained in the question (notably in the singular NP _which T-shirt_) that he bought one T-shirt. Simply answering the question would indicate that the answerer has no objections to this assumption. However, this assumption does not need to be contradicted verbally; a speaker can employ intonation to indicate that an assumption contained in a question is mistaken:

(44) Q Which T-shirt did you buy?
   A The red one AND the yellow one.

Thus we see that, apart from explicit verbal means, intonation (here in the form of heavy accent) may be employed as well to cancel an exhaustive interpretation.

The case of (43A1) has a special interest in the light of Horn's theory (1985, 1989) of metalinguistic negation. The negation in (43A1) is, like the negation in example (44A2) of section 3.5.2, unambiguously truth-functional, and thus truth-conditional: it states that the cardinality of the set of shirts bought by the speaker is not 'one' but 'two'. It is a simple and unmarked use of the negation operator, not requiring, though allowing for, heavy accent or stress. Yet this negation also clearly falls under the definition of metalinguistic negation as given in Horn (1985, 1989), which must be embarrassing for Horn's analysis.

According to Horn, there is just one negation operator, which is the one defined as a simple truth function in standard bivalent logic. Yet, he claims, the negation is 'pragmatically ambiguous' in that it can be used in two different ways. He distinguishes between an
unmarked, truth-functional use of negation, which he calls descriptive, and a marked non­truth-functional use of negation, which he calls metalinguistic:

While two distinct uses of sentential negation must indeed be admitted, the marked, nondescriptive variety is not a truth-functional or semantic operator on propositions, but rather an instance of the phenomenon of METALINGUISTIC NEGATION – a device for objecting to a previous utterance on any grounds whatever, including the conventional or conversational implicata it potentially induces, its morphology, its style or register, or its phonetic realization.

(1989:363)

Apart from the fact, pointed out by a number of authors, that it is difficult not to interpret the distinction between the two uses made by Horn as an straightforward form of logico-semantic ambiguity, we note here that the negation in (43A1), though truth-functional and 'descriptive' in Horn's sense, also clearly falls under his definition of metalinguistic negation. For it is "a device for objecting to a previous utterance", in this case on the grounds that speaker who is objecting regards the previous utterance as truth-conditionally false.

We may add that there is something that makes the negation in (43A1) special in some sense: it denies a presupposition of the previous utterance. The previous utterance (Which T-shirt did you buy?) presupposes that the addressee bought precisely one T-shirt, and in order to answer correctly the second speaker first has to correct that presupposition. This he does by saying "I didn't buy one, I bought two."

In Seuren (1995) it is argued that the class of cases dubbed 'metalinguistic' by Horn in fact falls into two distinct classes. One class involves ordinary unmarked and presupposition-preserving negation (minimal negation in his terminology). The other class involves a literal repetition of the presupposition-carrying utterance and implies that one or more of its presuppositions are false and must be corrected (radical negation, in his terminology). One distinguishing feature of the radical negation is that it can occur only in the 'canonical' position of negation, i.e. in construction with the finite verb of the sentence, and only in assertions, not in other sentence types. It must, moreover, be marked by heavy accent, as in (45a):

(45) a. She did NOT have dinner with the king of France: there is no king of France!
b. Not everybody had dinner with the king of France
c. !! Not everybody had dinner with the king of France: there is no king of France!

The minimal negation, on the contrary, may occur in many different positions in sentences of all types, as in (45b), where it occurs in front position. In fact, (45c) strikes one as incoherent, precisely because the negation there, owing to its front position, can only be the minimal negation and can thus not be used to imply the falsity of a presupposition of the original carrier sentence, i.e. (45a). On the basis of this and similar examples Seuren argues that the metalinguistic character of negation use, as defined by Horn, is no doubt real (and part of the general fact that metalinguistic comments are woven into ordinary language use in many different ways), but that, contrary to what Horn claims, the radical negation is a separate category within the wider category of metalinguistic negation, and that the remaining cases of metalinguistic negation are metalinguistic instances of the minimal (presupposition-preserving) negation.
In order to show that the negation in (43A1) is indeed Seuren's presupposition-preserving minimal negation we can apply the test that is used in (45c):

(46) a. Why did everyone buy a T-shirt?
   b. presupposition: "everybody bought a T-shirt"
   c. Not everybody bought a T-shirt. Some bought none.

(46a) carries the presupposition that everybody bought a T-shirt (i.e. (46b)), which is denied and corrected by (46c). The position of not in (46c) clearly shows that this is an instance of minimal, therefore unmarked, yet metalinguistic negation.

4.3.2.3 Correcting the exhaustivity
There is (at least) one argument against the analysis of exhaustive interpretation developed so far. It comes from multiple sentence answers (as J. Groenendijk (p.c.) pointed out to me):

(47) Q Who came to the meeting yesterday?
   A Ted came. Bill came, too.

If we analyze the two parts of (47A) separately as answers to the same question, thus as "Only Ted came, and only Bill came", then we have a contradictory result. At first sight it would seem that the theory developed here does exactly that. If we assume that every utterance is the answer to a question, then at first sight we have here two mutually inconsistent answers to the same question. However, this is not the correct analysis. The discourse in (47) is assumed to have the structure as in (48) (although the exact phrasing of Q2 is entirely speculative):

(48) Q1 Who came to the meeting yesterday?
   A1 Ted came.
   <Q2 Is that all?/Who else came?>
   A2 Bill came, too.

A2 thus is an answer to a question asking for more information, and it is hedged: the use of too here signals that the information in A2 is additional to the information already given in A1. In other words, the exhaustivity inherent in Q1/A1 is overruled by the use of too. The necessity of using too (or also, as we will see in a moment) is clear from the fact that omitting it leads to a piece of discourse which is hard to interpret properly:

(48) Q1 Who came to John's party last night?
   A1 Ted came.
   A2 Bill came.

With respect to this piece of discourse, there are at least two different interpretations possible, and intonation is essential in distinguishing them. (48') as a whole may be interpreted in two ways: either both Ted and Bill came to the meeting yesterday, or only Bill came (the speaker corrects himself). To support the interpretation that both Ted and Bill came to the meeting, the absence of a "terminal" (falling) intonation contour on Ted came is necessary. However, if both Ted came and Bill came are pronounced with a falling intonation contour on each phrase, the most likely interpretation of (48'A2) in the context given is that of a self-initiated self-
correction, i.e. not Ted, as was stated initially, but Bill came to the meeting. However, to be a clear and indisputable case of correction it should have been adequately marked as such. In general, and especially without any information with respect to intonation, it is hard to determine what a coherent interpretation of this piece of discourse would be.

A further example shows the same point. As we assume that exhaustive interpretation is related to questions, it is natural to assume also that it is answers which are interpreted exhaustively. And answers come in different forms, not only as single sentences. In fact, sentence boundaries can be rather arbitrary, as the difference between (49) and (50) shows:

(49) Q  How many men and women were present?
     A  There were two women, three men, and a sixth person present.62

(50) Q  How many men and women were present?
     A  There were two women and three men present. There was also a sixth person present.

(49A) in fact does not adequately answer (49Q); the speaker does not specify the numbers of men and women present, because he does not know the exact numbers. The same (inadequate) answer can be given in a different format, as in (50). The information structure of the latter piece of discourse is similar to that of (48) above, namely

(51) Q1 How many men and women were present?
     A1 There were two women and three men present.  
     <Q2 Is that all?>
     A2 There was also a sixth person present.

And here again we encounter a form of hedging one's answer in order to overrule the exhaustivity of an earlier answer: also is necessary in order to arrive at the desired interpretation. If also is omitted, the result is an uninterpretable piece of discourse.

4.3.2.4 Non-exhaustive questions

As Beck and Rullmann (1995) point out, not only answers, but questions can be marked explicitly for non-exhaustivity as well. In English, one could add for example in order to indicate that the question is not intended exhaustively:

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

Moreover, they point at the German so and the Dutch zoal, which seem to serve the same function of indicating that this particular (embedded) question does not require an exhaustive answer:

(53) Hans will wissen, wer so auf dem Fest war.

Hans wants to know who [so] at the party was

62 This example and the next one by themselves pose an interesting problem for implicature theory, as J. Groenendijk (p.c.) pointed out to me: as implicatures are derived on an item-by-item basis, the interpretation of this answer will inevitably run into a contradiction. Two women will get interpreted as "exactly two", three men as "exactly three", but these interpretations are not warranted because of the sixth person, whose sex is unknown.
"Hans wants to know who for example were at the party"

(54) Jan wil weten wie er zoal op het feest waren.
Jan wants to know who there [zoal] at the party were
"Jan wants to know who for example were at the party"

In this subsection (4.3.2) I have illustrated some ways in which speakers can prevent the exhaustivity inherent not only in explicit question-answer sequences, but in every piece of discourse. The claim is that exhaustivity can only be blocked explicitly, thus by using any of a variety of linguistic indications, including verbal means and intonation. To put it in classical pragmatic terms: exhaustive interpretation is a default process. An utterance is interpreted exhaustively, unless it is indicated explicitly that it should not be. From this description one can deduce that a default process in language can only be overruled explicitly. In other words, I claim that a process such as implicit cancellation, which Radical Pragmatics is forced to assume in order to account for the most elementary facts, does not exist at all.

The brief discussion given in this subsection is not intended as an exhaustive listing of all the linguistic means that speakers can employ in order to prevent the exhaustive interpretation of their utterances. Neither do I want to claim that the theoretical role that is attributed to the phrases discussed here, is the only one they can fulfill within a discourse. Problems like these are of a genuinely linguistic nature, and a much larger and more detailed survey of linguistic data than the one conducted here would be necessary in order to understand fully which linguistic means a speaker can use to prevent exhaustivity, and what roles the phrases cited here can play in a discourse. The only claim made here is that there are various strategies for a speaker to prevent the exhaustive interpretation of his or her utterance, and that each of these strategies can be realized linguistically in a variety of ways.

4.3.3 The exclusion inference

As we discussed in detail in chapter 2, one of the essential characteristics of the exclusion inference according to Radical Pragmatics is the fact that the generation of an exclusion inference is dependent on the availability of some expression alternative. If there is no expression alternative available, no Q1 implicature is generated. The theory presented here is structured differently: the exclusion inference is a type of inference that consists of two independent processes. The first of these is a process of exhaustive interpretation, which is fully integrated into the semantics of questions and answers, and for which the determining factor is the information structure of the discourse. Together with the assumption that all utterances in a discourse are the answer to a (mostly implicit) question, this leads to the prediction that in every utterance, some information (specifically the information obtaining the comment position in the information structure of the discourse at a particular moment) is interpreted exhaustively. The effect of exhaustive interpretation can be informally characterized as follows: an utterance $X$ which is interpreted exhaustively, is in fact interpreted as equivalent to only $X$. This interpretation thus has the immediate side-effect that any other possibilities are implicitly, but truth-conditionally discarded. The second process that makes up the exclusion inference is the (implicit) denial of those other possibilities.

My aim in this work has been to give an account of the first part of this process, but I wish to pay some attention to the second part also. In fact, most of the literature on implicatures is devoted to the problem of identifying the available expression alternatives,
which comes in various guises. The most familiar form is that of determining the necessary and sufficient conditions for Horn scales. As argued in chapter 2, this is only one particular version of the general problem that exists for the theory of Q1 implicature, which is to determine in any case, which are the available expression alternatives. As I put it before, informally, when a statement is made, something is affirmed, but simultaneously something else is (implicitly, but truth-conditionally) denied. This is the inference that I named the exclusion inference, and that in Radical Pragmatics was identified as (a particular form of) generalized Q1 implicature. This observation is also made by De Morgan (1847:4, also cited in Horn 1989:211):

In common conversation the affirmation of a part is meant to imply the denial of the remainder. Thus, by 'some of the apples are ripe', it is always intended to signify that some are not ripe.

The problem is, what is this "remainder" that is being denied? In this section, I will not review the solutions proposed within implicature theory, nor will I propose a solution of my own. On the contrary, I believe this to be an entirely separate issue, which falls outside the scope of the present study. Nevertheless some brief remarks are in order here. First I will show how the denial of any available expression alternatives follows from the process of exhaustive interpretation as it was defined here, and secondly I will briefly discuss some issues which are related to the problem of the available expression alternatives in various ways.

The exclusion inference follows quite easily from the exhaustive interpretation of (parts of) an utterance. This process can be described informally as follows: in discourse, there is always some part of an utterance that is interpreted exhaustively. This means that this part is in fact interpreted more strongly than it was uttered. Thus, an expression \( X \) in comment position is interpreted to mean "\( X \) and nothing but \( X \)", "\( X \) and nothing else". In other words, "the denial of the remainder" is an immediate consequence of this interpretation. In the theory presented here, this is integrated into the semantics of \( \text{besp} \) as follows. In section 3.2.1 the semantics of \( \text{besp} \) is defined as \( \{<a,b> | a \text{ is an argument for a function } f, b \text{ is its value}\} \). As was noted there, it is crucial that the relation between \( a \) and \( b \) is a function, i.e. that for any argument \( a \) only one value \( b \) is selected from the co-domain. Thus, if the function is the one who hit Mary, with as argument a specific discourse-determined situation, and the value (specification) given is John (as in (55)),

(55) Q Who hit Mary?
A John did.

then, by the fact that \( \text{besp} \) is a function, all other possible values from the co-domain are excluded (implicitly denied). The selection of one value from the co-domain implies the denial of all other values from the co-domain. To put it in set-theoretic terms, from the affirmation of an element \( b \) from a set (i.e., the co-domain) as value for a particular function \( f \) and a particular argument \( a \) follows the simultaneous denial of all other elements from that same set, i.e. the denial of the complement of \( b \), as correct values for the function \( f \) and the argument \( a \). That this is a correct characterization of the exclusion inference in set theoretic terms is probably most clearly seen from the case of the restrictive subclauses (example (27), repeated here):

(56) Q Who got lost in the forest?
A The men who were on their way home (got lost in the forest).

The exclusion inference induced by (56A) is that the men who were not on their way home, did not get lost in the forest. This inference is formed as follows. The co-domain here is the set of men, from which the uttering of (56A) selects the subset of those who were on their way home as the value for the function getting lost in the forest and the argument of a specific discourse-determined situation. The complement is the subset of men who were not on their way home, and of this complement it is denied that they form a correct value for the function getting lost in the forest in the indicated discourse-determined situation: thus, the exclusion inference is that "the men who were not on their way home did not get lost in the forest."

Other examples show the same inference type, though maybe not as clearly as in this particular case. In (57), the function required by besp is $b$ is the colour of $a$ and the argument $a$ is identified by the phrase that flag. The relevant co-domain is that of the colours.

(57) Q What is the colour of that flag?
A It is white.

The value selected from the co-domain by (57A) is white, and its complement comprises all colours that are not white, thus red, blue, green, yellow etc. etc. The exclusion inference now is the denial of this complement: "the colour of that flag is not (not white), thus not red, not blue, not green etc." The example in (58) involves one of Horn's scalar predicates. Here, the function is $b$ is the temperature of $a$, and the argument $a$ is the water (which is unambiguously identified by other aspects of the discourse-determined situation). The relevant co-domain is that of temperatures.

(58) Two persons are going for a swim in the lake. One has already tried it, the other hasn't.
Q What is the temperature of the water?
A It is cold.

By selecting the value cold from the co-domain, the speaker implicitly denies all other values in the co-domain, including a value such as cool. This is different from Horn's account of this particular case: on Horn's account, the selection of the value cold from the scale <cool, cold, freezing> results in the denial of only the stronger value freezing, as a result of the fact that the concepts on this scale are ordered by one-sided entailment.

In the theory presented here, the question so prominent in the theory of Q1 implicatures, namely "what are the available expression alternatives?" is equivalent to the question "what is the relevant co-domain for the function $f$ and the argument $a$?" Obviously, this by itself does not solve the problem; however, the gain is that the role of the set of alternatives is made very clear. The theoretical problem of determining the relevant co-domain for a given utterance is an extremely hard one, to whose solution I can hardly hope to contribute in this study. However, I can point out that the theory defended here is consistent with Hirschberg's observations with respect to, in her terminology, the POSETS that license a Q1 implicature. POSETS are primarily conceptual orderings; partial orderings over expressions are a reflection of these conceptual orderings:

Posets supporting scalar implicature may be defined over (expressions that denote) entities, actions, attributes, times, places, or concepts, including those concepts
ordered in Horn's canonical quantifier, modal, and number scales – or any other
items. (Hirschberg 1991:125)

One could imagine a theory in which Hirschberg's POSETS occupy the place of the co-domain
demanded by the semantics of the abstract predicate besp. However, the details of such a
type I leave to a later date.

So far in this study, we have silently adopted a limitation which in fact is not justified.
We have only considered affirmative statements functioning as answers, and similarly only
considered non-negative questions. However, no special arrangements are needed in order to
account for negative questions or negative answers. The theory developed here accounts for
these cases as well. In set-theoretic terms, the affirmation of a value b for a function f and a
particular argument a leads to the implicit denial of the complement of b within the co-
domain; and similarly, the negation of a value b for a function f and a particular argument a
leads to the implicit affirmation of the complement of b within the co-domain. This is an
example involving a negative question:

(59) Sophie is eating a peach.
    S. I can't eat this.
    R. What is it you can't eat?
    S. The stone.

From the final utterance it can be concluded that only the stone of the peach isn't edible, or in
other words that the whole peach except its stone is edible.

I restated the problem of determining the available expression alternatives as
determining the relevant co-domain for a certain function f and a value a. In restating it thus,
I used the word relevant deliberately. The main reason for this is that I think that finding an
adequate general characterization of the notion of relevance, which plays such an important
role in implicature theory, will largely solve this problem. Finally, it is important to note that
exhaustive interpretation is a process that is totally independent of the availability of an
expression alternative. Exhaustive interpretation is a structurally determined process. The
generation of exclusion inferences follows from this immediately, but only when it is clear
somehow (e.g., it is indicated or stated in the previous context) that there is a co-domain
available with more than one member. If the co-domain has only one member, i.e. the one
specified as value for the function f and the argument a, then this is taken as an exhaustive
specification without any further consequences (i.e., without an exclusion inference
following).

4.4 Exhaustive interpretation as a form of Q1 implicature

The possibility of a relation between the process of exhaustive interpretation and
conversational implicature has been pointed out before. In their dissertation (1984b),
Groenendijk and Stokhof (henceforth G&S) develop a theory of the semantics and pragmatics
of questions and answers, within the theoretical context of formal semantics. They consider it
characteristic of linguistic answers that they are exhaustive; and the observations that they
make with respect to exhaustive interpretation are entirely consistent with the observations
made in the present study. G&S characterize exhaustiveness as follows: (60A), when
functioning as answer to (60Q), has a different interpretation than when interpreted in isolation.

(60) Q Who are walking in the garden?
A John and Mary are walking in the garden.

(61) Only John and Mary are walking in the garden.

Specifically, (60A) as answer to (60Q) is equivalent to (61), whereas in isolation it is not. Or in other words, "taken in isolation the interpretation of the indicative sentence in [60A] is such that its truth is compatible with other people walking in the garden as well." (274) Thus, exhaustiveness is regarded as a property of answers to questions, which moreover has truth-conditional effects. These truth-conditional effects are formulated differently as well. With respect to (62A1) and (62A2)

(62) Q Who is walking?
A1 John is walking.
A2 John and Mary are walking.

it is observed that

... although taken in isolation, the truth of [62A2] implies the truth of [62A1], as answers to [62Q] they contradict each other. (295)

A further property that G&S immediately note is the default-nature of exhaustiveness: an answer is exhaustive unless it is (explicitly) indicated that it is not. These observations with respect to the nature of exhaustiveness, and specifically the observation that it is what I have called a relational notion, for G&S form a conclusive argument against the possibility of interpreting a sentential\textsuperscript{63} answer independently of the question to which it functions as answer.

Formally, exhaustivization is introduced as a separate semantic operation on the constituents (terms) from which the answers to questions are derived. G&S are however not very happy with this solution, especially not with the semantic character of exhaustivization. In an appendix to the chapter devoted to the semantics of questions, they throw up the question whether exhaustiveness could be a form of conversational implicature:

Why, one may ask, isn't exhaustiveness simply obtained as a conversational implicature? If anything is a good candidate for implicaturehood, exhaustiveness of answers is, or so it seems. (368)

The reason for preferring a pragmatic explanation for exhaustiveness over a semantic one lies in the default character of exhaustive interpretation: it appears unless it is explicitly cancelled,

\textsuperscript{63} G&S distinguish between what they call sentential (A1) and constituent (A2) answers. Of members of the latter category it is obvious that they should be related to a question in order to arrive at the correct interpretation; of members of the former category, this is less obvious however.

Q Who is walking in the garden?
A1 John and Mary are walking in the garden.
A2 John and Mary.
just as Gricean conversational implicatures. However, in order to formulate an adequate pragmatic (Gricean) explanation for the process of exhaustive interpretation, a "formal statement of the requirements inherent in the Gricean Maxims" (369) is necessary. They argue that their own formalization of the Maxims, as given in their (1984a), will not do for the purpose of giving a pragmatic explanation of exhaustive interpretation.

The line of argument that G&S propose in this appendix is in fact the opposite of the line I am pursuing here: they suggest that the process of exhaustive interpretation is better explained, not as a semantic operation, but as a form of conversational implicature, while I have been arguing the claim that what implicature theory regards as (non-truth-conditional) Q1 implicature is in fact the (truth-conditional) result of a process of exhaustive interpretation. In all it may be said that, apart from the formal details, which I will not go into any further here, the characterizations of exhaustive interpretation given here and that of G&S are consistent. The chief difference between the present study and that of G&S is that the latter is intended as a contribution to the (formal semantic) theory of questions and answers, while my discussion is with implicature theory.

4.5 Strong vs. weak theorists

In much of the relevant literature, Q1 implicature is depicted as the only alternative to entailment as an explanation of the type of inference at issue. In other words, a dilemma is posed: strong theorists are those that maintain that the exclusion inference from (63a) to (63b) is a case of entailment, weak theorists maintain that it is some form of conversational implicature.64

(63) a. The flag is red.
   b. The flag is all/ only red.

The aim of the discussion in this section is to show that this dilemma is a false one. It can be circumvented if the context of an utterance is systematically taken into account. I will first discuss Harnish's (1976) arguments to regard the exclusion inference induced by (63a) as a form of generalized conversational implicature; my main interest is in the dilemma he assumes and the arguments he puts forward to make a choice in this dilemma. Next, I will discuss Horn's (1989) position with respect to the semantics of scalar terms, which is the result of positing a similar (and equally false) dilemma.

In a paper aimed at identifying and distinguishing various forms of inference in natural language (entailment vs. implicature, entailment vs. presupposition), Harnish (1976) proposes a modification to Gricean implicature theory in order to deal with the exclusion inference from (63a) to (63b).65 He starts by arguing that the exclusion inference cannot be a case of

---

64 This use of the terms strong and weak theorists is adopted from Harnish, who borrows the terms (with a slight change in meaning) from Grice.

65 Harnish in fact rests his discussion on two examples, namely (63) above and (i):

(i) Russell and Whitehead wrote Principia.

(i) is taken to induce the inference "Russell and Whitehead wrote Principia together". This example involves the problem of singular and group readings, which is a complicated matter. As it lies entirely outside the topic of this discussion, I will not discuss this example.
entailment. He presents two types of arguments to support this claim: the first type of argument is to show that the strong theorists' position leads to a contradiction, the second type of argument is to show that the exclusion inference has properties which rule out the possibility that it is entailment. The claim that the strong theorists' position leads to a contradiction is supported by two arguments. First, supposing that (64) is true,

(64) The flag is red and white.

then for semantic reasons it is not possible (on the strong theorists' position) to derive\textsuperscript{66} this reduced conjunction from the separate conjuncts in (65):

(65) a. The flag is red.
    b. The flag is white.

The strong theorist claims that (65a) and (b) entail (66a) and (b):

(66) a. The flag is all (only) red.
    b. The flag is all (only) white.

However, given that (64) is true, (66a) and (b) are false and thus, by contraposition, (65a) and (b) as well. Thus the strong theorist would be making the inconsistent claim that the true sentence in (64) is derived from the false sentences in (65). Harnish's own position is, first, that "transformations do not change meaning" (315, 320), and secondly, that the sentences in (65) are true in the given context (i.e., if (64) is true) (315, 320). Thus, if one wants to save the syntactic derivation from (65) to (64) then it is necessary to challenge the relationship of entailment between (65a) and (66a), and between (65b) and (66b).\textsuperscript{67}

Harnish's second argument for the fact that the strong theorists' position leads to a contradiction is more relevant to the topic of this dissertation. Again, suppose that (64), repeated here as (68), is true. Then

\begin{align*}
\text{suppose someone asks} & \\
[67] \text{What colour is the flag?} & \\
\text{and I answer} & \\
[68] \text{(The flag is) red and white.} & \\
\text{But if someone asks [67] and I answer,} & \\
[69] \text{(The flag is) red.} & \\
\text{then [69] entails that} & \\
[70] \text{(The flag is) all (only?) red.} & \\
\text{Since [70] is inconsistent with [68] and [68] is (supposedly) true, [69] which entails [70] must be false. (320/1)} &
\end{align*}

Harnish objects to this conclusion, because he maintains that (69) is true whenever (68) is true (thus that (68) entails (69)). What is remarkable about this argument is the fact that Harnish introduces an explicit question as context, but does not incorporate the role of this context into his conclusion. He concludes that the strong theory, which claims that the relation

\textsuperscript{66} Harnish presumably means: derive syntactically, although he does not state this explicitly.
\textsuperscript{67} Harnish's aim with this argument is not to save the operation of conjunction reduction, but only to exhibit the falsity of the premiss that (4a) entails (5a) and (4b) entails (5b) (see his conclusions on p. 364).
between (68) and (69) is entailment, is misguided. Harnish's claim, made in this and the
previous argument, could be summarized as follows: it is not possible to maintain both that
(71) entails (72a) and (72b), and that (72a) and (b) entail (73a) and (b), respectively:

(71) The flag is red and white.

(72) a. The flag is red.
   b. The flag is white.

(73) a. The flag is all (only) red.
   b. The flag is all (only) white.

However, if we incorporate the context into this discussion, and assume that both (71) and
(72a,b) are answers to the question *What colour is the flag?*, then in this context (71) does not
entail (72a) and (72b). For the information structure of these examples is such that in (71), the
expression *red and white* is in comment position, in (72a) *red*, and in (72b) *white*. These are
each interpreted exhaustively, which excludes the possibility of a relation of entailment
between (71) on the one hand and (72a,b) on the other hand. However, in this context, the
latter two do entail (73a) and (73b), respectively.68

Harnish's second type of argument concerns the characteristics of the exclusion
inference. His argument is that the inference from (67) to (68) can be cancelled, i.e. that it can be
denied without a contradiction arising:

If P entails Q, then it cannot be the case that \( P \text{ and not } Q \) is true; \( P \text{ and not } Q \) is a
contradiction. (...) Suppose that

[74] a. the flag is red entails the flag is all red.
then it should be the case that
[74] b. the flag is red, but it is not all red.
is a contradiction. But is does not seem to be, so [72a] is false and not a case of
entailment.69 (323)

On the basis of these two types of argument Harnish concludes that the exclusion
inference is not a case of entailment. In this conclusion, the role of context is not

68 This view of the relationship between (71), (72) and (73) in fact excludes the possibility that (71) is
(syntactically) generated by conjunction reduction (as understood by Harnish) from (72a) and (72b). The
utterance in (71), as answer to *What colour is the flag?* is both semantically and syntactically independent from
the utterances in (72) as answers to the same question. In this respect, (71) in the given context is comparable
rather to a conjunction such as (i), than to a conjunction such as (ii):

(i) Sue and Harry are a nice couple.
(ii) John and Bill live in London

The form of constituent conjunction illustrated in (i) can never be derived by conjunction reduction from the phrases *Sue is a nice couple* and *Harry is a nice couple*. (ii), on the other hand, could in principle be derived by conjunction reduction from *John lives in London and Bill lives in London*. For a discussion of conjunction
reduction, see Van Oorsouw (1987).

69 In this quote, Harnish uses the term *contradiction* incorrectly. When two sentences P and Q are contradictory,
then the truth of P follows from the falsity of Q and vice versa. This is too strong a claim in this case: if *the flag
is all red* is false, then it does not follow that *the flag is red* is true (for it might be any other colour). In both
instances, the correct term is *inconsistency*. 
incorporated. Harnish proposes an alternative account in terms of Q1 implicature, the details of which are not relevant to the present discussion.

Harnish's arguments discussed so far only concerned examples where the lexical material inducing the exclusion inference was in comment position. He also (accidentally) gives an example where the lexical material inducing the exclusion inference is in non-comment position. Harnish constructs his initial Gricean arguments for the inference from the flag is red to "the flag is all (only) red" in the context of the question What colour is the flag? Next he observes that

(...) one advantage of treating "the flag is all red" as a conversational implicature and not as an entailment is that in a conversation requiring the identification of a certain colour, one can point to an American flag and say that flag is magenta without implicating that it is all red. In this case, the implicature is contextually cancelled. (357)

Obviously, Harnish's observations here are entirely consistent with the view developed here, but his explanation in terms of contextual cancellation is not. From his description of the context we can conclude that the colour-term magenta would be in non-comment position in this discourse:

(75) Q What object has the colour magenta?
    What does magenta look like?
    A That flag is magenta.

Thus, there is no notion of implicit cancellation necessary to account for the fact that from this utterance, the hearer will not infer that the flag is only magenta and not some other colour. Again in this example, Harnish invokes the context, but does not draw the conclusion relative to this context.

Back to the strong and weak theorists: I believe to have shown in the previous sections, and indicated again in the discussion of Harnish's arguments, that there is a way out of this dilemma without choosing sides between either a strong or a weak position. If the context is systematically taken into account, the various readings that Harnish identifies for his examples (notably (75A) in the context specified, and (68) as an answer to (67)) are easily accounted for. As a result, the exclusion inference is strong (i.e., entailment) when it is invoked by lexical material in comment position, but weak (invited inference) when invoked by lexical material in non-comment position.

Horn (1989) poses a dilemma similar to that of Harnish, between what he calls "unilateralists" and "radical bilaterals". As we saw in detail in section 2.1.5, the Radical Pragmatics version of implicature theory is based on a particular view of the semantics of scalar predicates, i.e. those lexical items which are assumed to induce a scalar implicature. The semantics of e.g. some is "at least some". Or, in Horn's words: some "asserts lower boundedness". Horn argues for this (intuitively unattractive) position by pointing out that an alternative analysis has unwanted consequences:

If some really meant, rather than implicating, 'some only', some at least would be a contradiction. [but see note 14, R.S.] Furthermore, the radical bilateralist would predict that expressions like some but not all, and indeed some only or only some
(along with the examples of [76]) would be semantically redundant, which they clearly are not.

[76] reinforcement

Pat has exactly three children
three and only three children
three children but not four

Horn (1989:215)

As Harms did in his (1976), Horn discusses only two options here, namely being "unilateralist" (some means "at least some") and "radical bilateralist" (some means "some but not all"). But here again, it is not necessary to choose either of these options; it is possible to choose a middle way and defend what one could call a "moderate bilateralist" position.

There are three central aspects of the moderate bilateralist position. First of all, some means "some"; in the semantic paraphrase of any lexical item, no "at least" is necessary. Secondly, the interpretation of lexical items is in part determined by their position in the information structure of the utterance they are part of; notably, items in comment position receive an exhaustive interpretation ("only some"). And thirdly, as a consequence of this, and on the assumption that semantic redundancy should be avoided in language, the constructions that Horn lists as semantically redundant could in this view be distributionally constrained to a particular place in the information structure of the text. E.g., with respect to the numerals, exactly n would preferably be in non-comment position (as it is synonymous with n in comment-position), while at least n would preferably be in comment-position. This is a hypothesis which can easily be tested. However, the details of this I leave for further investigation.

4.6 Summary and conclusion

In chapter 1 we saw how Horn adopts Grice's philosophical theory of conversational implicature as the solution to a complex of problems. This complex includes, first, the so-called one- and two-sided readings of various predicates, secondly, the characteristic "less than" interpretation of numerals under negation, and finally the nature of the relationship between the subcontrary statements some men are mortal and not all men are mortal. All these problems are simultaneously solved, by Horn, by the adoption of generalized conversational implicatures induced by Grice's first submaxim of Quantity, which reads "Make your contribution as informative as is required (for the current purposes of the exchange)". The theoretical context developed in this book is chiefly aimed at presenting an alternative solution for the first of these problems; an alternative solution to the second problem is shown to follow easily from the theory developed in chapter 3.

The alternative account for the one- and two-sided readings of various predicates presented here differs in some important aspects from the account in terms of implicature. Most previous accounts (including the account in terms of generalized conversational implicature) build on the implicit assumption that this problem is one of lexical semantics and pragmatics. There are at least three such proposals. First, it has been suggested that the predicates involved are either ambiguous between the one- and the two-sided readings, or secondly that the predicates involved have vague meanings, which leave considerable room for interpretation. And finally, it has been suggested that the predicates have the one-sided
meaning as their literal meaning, and that the two-sided interpretation is the result of a conversational implicature. However, the alternative presented here says that the problem of the one-and two-sided readings is a problem related to the information structure in discourse, specifically, that the meaning, and the truth conditions, of a string of words (a sentence) varies with the context in which it functions (specifically, the question to which it functions as an answer).

In order to make the development of an alternative for Horn's theory of scalar (Q1) implicatures possible, it was necessary to take some preliminary steps. The first was to rename the inference characteristically referred to as Q1 implicature, in order to remove the theoretical association with implicature theory. Based on an intuitive characterization we introduced the term exclusion inference. The second step was the adoption of Discourse Topic Theory (DTT) as a theoretical account of the information structure in discourse. DTT analyzes the terms topic and comment in terms of questions and answers. Every (explicit or implicit) question constitutes a topic, the comment is specified by the answer. In this way, DTT provided us with a means of systematically taking the context of an utterance into account: every utterance is regarded as the answer to a (implicit or explicit) question. This move has a clear methodological advantage: it makes it possible to incorporate the role of context systematically into the discussion.

The discussion in chapter 2 centered around (one particular version of) the so-called Gricean argument. This argument is considered by Grice, and by his followers in the tradition of Radical Pragmatics, to be one of the main indicators of the presence of a conversational implicature. The formulation of the Gricean argument shows up very clearly the two fundamental premises of implicature theory which we challenged in this study: first of all the assumption that scalar predicates are, as Horn puts it, "lower-bounded by their literal or conventional meaning", thus that e.g. a numeral $n$ means "at least $n$"; and secondly the assumption that the exclusion inference is dependent, for its inducement, on the availability of some expression alternative. If there is no expression alternative, then no exclusion inference is induced.

The opposing hypotheses defended here are as follows. The assumption that scalar predicates have an "at least"-meaning is abandoned completely. As we show in detail in chapter 3, with respect to the numerals, it is possible to account for the one-and two-sided readings in a different way. If we take into account the place that a particular predicate occupies within the information structure of the discourse, then it is possible to account for both the "at least" and the "exactly" readings of numerals under the assumption of a unitary, "exactly"-semantics. Specifically, if a numeral is in comment position (when it is in a strict sense the answer to a question) it receives the "exactly" or cardinality reading. When it is in non-comment position it receives the "at least" or existential reading, due, not to an inherent "at least" reading of the numeral in question, but to the existential quantifier. The second assumption, that the inducement of an exclusion inference is dependent on the availability of an expression alternative, is replaced by the following hypothesis: in the generation of an exclusion inference, two independent processes are operative. The first is a process of exhaustive interpretation, which is fully integrated into the semantics of questions and answers. Whether some specific piece of information is interpreted exhaustively is not in any way dependent on the characteristics of the concepts expressed or the lexical items used, but is determined by the place that the expressions occupy in the information structure of the discourse. Specifically, lexical material in comment position is interpreted exhaustively. The effect of exhaustive interpretation is that the interpretation of some expression $X$ is strengthened to "only $X$, and nothing else". Thus, by the exhaustive assertion of $X$ any
available expression alternatives are intrinsically, but truth-conditionally, denied. The second process operative in the generation of the exclusion inference is the selection of the relevant set of expression alternatives; this is a complex problem which is far from solved.

The details of the process of exhaustive interpretation are discussed in chapters 3 and 4. In chapter 3, the technical aspects of the semantics of questions and answers are summarily discussed within the theoretical context of discourse semantics. Exhaustive interpretation is not a separate operation on questions or answers, but is integrated into the common structures underlying questions and answers. In this common structure, an abstract predicate \( b_{sp} \) is used, which has the following satisfaction conditions: \( [[b_{sp}]] = \{<a,b>| a \text{ is an argument for a function } f, b \text{ is its value}\} \). In other words, \( b_{sp} \) requires a function \( f \) relating an argument and a value, and this is crucial: by definition, any function relates an element from its domain to exactly one element from its co-domain. Consequently, it is inherently asserted in the specification of a value, i.e. the selection of an element from the co-domain, that this is the only valid specification. In other words, the exhaustivity of such a specification is intrinsically asserted. But as is pointed out in chapter 4, in spite of this systematic occurrence of exhaustive interpretation, on the surface it appears as a default process. First of all there seem to be pragmatic principles limiting the applicability of exhaustivity, although these pragmatic principles were not considered in any detail. But more importantly, there are various strategies for speakers to indicate that either their (explicit) question or their answer be interpreted as incomplete, or doubtful, or anyway not simply exhaustive. These strategies can be realized by several different linguistic means, lexically or by the use of intonation. However, for a full account of the strategies available to relativize exhaustivity, much more detailed linguistic research is necessary, the examples and considerations presented here can only serve as a first indication.

The evidence presented to support this analysis can be found in chapters 3 and 4. In chapter 3, the numerals are discussed separately for reasons of exposition only. In general, for mutually exclusive predicates, among which are the numerals, it is established in chapters 3 and 4 that the relationship between interpretation and information structure is as follows: when such a predicate is in comment position (i.e., when it is in a strict sense the answer to a question), it receives the "exactly" or two-sided interpretation. We say that it is interpreted exhaustively. This is supported by the observation that in these constellations, cancellation by one of the well-known cancellation frames of the exclusion inference leads to an unacceptable piece of discourse. In other words, the exclusion inference is truth-conditional in such a constellation. On the other hand, when the predicate is in non-comment position it does not have to be interpreted exhaustively and correction is easy (although it necessarily involves a topic shift). For the classical cases of scalar predicates, the investigation carried out in chapter 4 yields slightly different results. Exhaustive interpretation of a scalar predicate like beautiful does not lead to the impossibility of a correction or cancellation. Such a correction however does have a rhetorical flavor to it. Correction of scalar predications in non-comment position goes through easily. The examples studied in the early sections of chapter 4 do support the conclusion that the scalar predicates are, at least with respect to their semantics, different from the mutually exclusive predicates.

The indications for further research that emerge from the present study chiefly point in the direction of linguistic research. If one accepts a principle such as Horn introduces, namely that semantic redundancy is avoided in language, and one moreover accepts the effects of discourse structure on the interpretation of utterances as presented here, then one can imagine two types of linguistic effects. The first of these is distributional effects in the following way: any phrase which has by itself an exhaustive meaning (examples are expressions such as...
"exactly n" and phrases with *only*) should not occur in comment position, as this would lead to a double exhaustive effect. However, a much more likely finding seems to be that the differences in interpretation between lexical material in comment position and in non-comment position should carry over to a much larger category of cases than was studied here. An interesting case would be the interpretation of an exhaustivity marker such as *only* in various structural positions; the theory presented here would predict a difference between its interpretation in comment position and in non-comment position. Questions such as these are left for further research.
bibliography


De Morgan, A. (1847), *Formal Logic or the Calculus of Inference, Necessary and Probable*. London: Taylor and Walton


**samenvatting**

In hoofdstuk 1 wordt uiteengezet dat Horn de filosofische theorie van Grice met betrekking tot conversationele implicaturen ziet als de oplossing voor een heel complex van problemen. Dit complex omvat ten eerste de zogenaamde één- en twee-zijdige lezingen van diverse predicaten; ten tweede de kenmerkende "minder dan" interpretatie van getalnamen onder negatie; en ten derde de aard van de relatie tussen de uitspraken *sommige mensen zijn sterfelijk* en *nieu alle mensen zijn sterfelijk*, die in het vierkant van Aristoteles de subcontraire uitspraken vormen. Deze drie problemen worden door Horn in één keer opgelost door de aanname van gegeneraliseerde conversationele implicaturen, geïnduceerd door het eerste submaxime van Quantiteit (kortweg Q1): "Maak je bijdrage zo informatief als nodig is (voor de actuele doeleinden van de conversatie)". De theoretische context die in dit werk wordt ontwikkeld is met name gericht op het geven van een alternatieve verklaring voor het eerstgenoemde probleem, dat van de één- en twee-zijdige lezing van predicaten: hoe komt het dat sommige predicaten, en de getalnamen zijn een voornam categorie in dit verband, in sommige gevallen moeten worden geïnterpreteerd als "tenminste" (één-zijdig), en in andere gevallen als "precies" (twee-zijdig). In hoofdstuk 3 wordt daarnaast kort aangegeven hoe een alternatieve verklaring voor het tweede probleem, de voorkeursinterpretatie van getalnamen onder negatie, als vanzelf volgt uit de daar ontwikkelde theorie.

Het alternatief dat hier wordt gepresenteerd met betrekking tot de één- en twee-zijdige lezing van predicaten verschilt in een aantal belangrijke opzichten van de verklaring in termen van conversationele implicaturen. De meeste verklaringen die tot dusverre voor dit probleem zijn gepresenteerd, en waarvan de theorie van de gegeneraliseerde conversationele implicaturen er één is, gaan uit van de impliciete aanname dat het moet worden opgelost op het niveau van de lexicaal semantiek en pragmatiek. Er zijn tenminste drie voorstellen die hiervan uitgaan. Ten eerste is voorgesteld dat de betrokken predicaten (lexicaal) ambig zijn tussen de één- en de twee-zijdige lezing. Ten tweede is voorgesteld dat deze predicaten een vage betekenis hebben, die aanzienlijke ruimte voor interpretatie laat. En tenslotte is voorgesteld dat deze predicaten de één-zijdige interpretatie als letterlijke betekenis (semantiek) hebben, en dat de twee-zijdige interpretatie een gevolg is van een conversationele implicatuur die door het gebruik van het predicaat wordt opgeroepen. Het alternatief dat hier wordt gepresenteerd gaat er echter van uit dat het probleem van de één- en twee-zijdige lezing van predicaten een probleem is dat niet op het niveau van de lexicaal semantiek en pragmatiek moet worden bezien, maar integendeel een probleem is dat samenhangt met de informatiestructuur in discourse. De claim is dat de betekenis, en de waarheidscondities, van een reeks woorden (een zin) varieert met de context waarin de zin functioneert (waarbij de context wordt gezien als de WH-vraag waarop de zin als antwoord fungeert). Het deel van de zin dat in strikte zin antwoord op de vraag is wordt exhaustief of uitputtend geïnterpreteerd, dat wil zeggen dat het de "precies"- of twee-zijdige interpretatie ontvangt. Het deel van de zin dat niet als antwoord op de vraag (in strikte zin) fungeert ontvagt de non-exhaustieve, "tenminste"- of één-zijdige interpretatie.

Om de ontwikkeling van dit alternatief voor Horn's theorie van scalaire (Q1) implicaturen mogelijk te maken is het nodig om een aantal voorbereidende stappen te nemen. Deze worden uiteengezet in paragraaf 1.4. De eerste stap is om de inferentie die algemeen wordt aangeduid als Q1 implicatuur een andere naam te geven, om de associatie met de theorie van conversationele implicaturen weg te nemen. Op basis van een intuïtieve karakterisatie van de inferentie wordt gekozen voor de term *exclusion inference*. De tweede
stap is de aannemer van Discourse Topic Theorie (DTT) als theoretische verklaring van de informatiestructuur in discourse. DTT analyseert de termen *topic* en *comment* in termen van vragen en antwoorden. Iedere (expliciete of impliciete) vraag constitueert een topic, de comment wordt gespecificeerd door het antwoord. Op deze manier stelt DTT ons in staat om de context van een uiting op een systematische manier te integreren in de bestudering van de interpretatie van die uiting: iedere uiting wordt beschouwd als het antwoord op een (impliciete of expliciete) vraag. Deze stap heeft een duidelijk methodologisch voordeel: de rol van de context op de interpretatie van een uiting wordt op een systematische manier opgenomen in de discussie.

De discussie in hoofdstuk 2 is opgezet rond (een specifieke versie van) het zogenaamde Griceaanse argument. De mogelijkheid om een dergelijk argument te construeren wordt door Grice, en ook door zijn navolgers binnen de traditie van de Radical Pragmatics, opgevat als één van de belangrijkste aanwijzingen voor de aanwezigheid van een conversationele implicatuur. De formulering van een dergelijk argument laat duidelijk de twee fundamentele aannames zien waarop implicatuurtheorie berust, en die in deze studie worden losgelaten: ten eerste de aannemer dat scalaire predicaat letterlijk een "tenminste" betekenis hebben, en ten tweede de aannemer dat het optreden van een exclusion inference of gegeneraliseerde Q1 implicatuur afhankelijk is van de beschikbaarheid van een alternatieve expressie (die aan bepaalde voorwaarden moet voldoen). Als er geen alternatieve expressie beschikbaar is, dan treedt er geen exclusion inference op.

De alternatieve hypothesen die in deze studie worden verdedigd zijn als volgt. De aannamer dat scalaire predicaat een "tenminste" betekenis hebben wordt geheel verworpen. Zoals in hoofdstuk 3 in detail uiteen wordt gezet met betrekking tot de getalnamen, is het mogelijk om de één- en tweezijdige lezingen op een hele andere manier te verklaren. Als de plaats die een getalnaam inneemt in de informatiestructuur van de discourse mee in beschouwing wordt genomen, is het mogelijk om zowel de "tenminste"- als de "precies"-interpretatie van die getalnaam te verantwoorden onder aannemer van een unitaire, "precies"-semantiek. De rol van de informatiestructuur is als volgt: als een getalnaam zich in comment positie bevindt (als hij dus in strikte zin fungeert als antwoord op een vraag) onvangt hij de "precies"- of cardinaliteitslezing. Als een getalnaam zich in non-comment positie bevindt ontvangt hij de "tenminste" of existentiële lezing, niet als gevolg van een "tenminste" semantiek, maar als gevolg van de aanwezigheid van de existentiële kwantor als hoogste predicaat in de semantisch-syntactische diepestructuur van de uiting.

De tweede aannemer karakteristiek voor implicatuurtheorie, dat het optreden van een exclusion inference afhankelijk is van de beschikbaarheid van een alternatieve expressie, wordt eveneens losgelaten. De alternatieve hypothese houdt in dat het optreden van een exclusion inference wordt bepaald door twee, van elkaar onafhankelijke, processen. Het eerste is een proces van exhaustieve of uitputtende interpretatie, dat geheel is geïntegreerd in de semantiek van vragen en antwoorden. De kwestie of een bepaald gedeelte van een uiting exhaustief wordt geïnterpreteerd is op geen enkele manier afhankelijk van eigenschappen van de gebruikte concepten of lexicale items, maar wordt beslist door de plaats die de expressies innemen in de informatiestructuur van de discourse. Alles wat in comment positie staat wordt exhaustief geïnterpreteerd. Het effect van exhaustieve interpretatie is dat de interpretatie van een expressie X wordt versterkt tot, informeel uitgedrukt, "alleen X, en niets anders". Aldus worden door de exhaustieve bewering van X alle beschikbare alternatieven intrinsiek, en waarheidsconditioneel, genegeerd of ontkend. Het tweede proces dat een rol speelt in de generatie van exclusion inferences bestaat uit de bepaling van de relevante verzameling van
alternatieve expressies; dit is een buitengewoon complex probleem, waar in deze studie niet verder op ingegaan wordt.

De details van het proces van exhaustieve interpretatie worden besproken in de hoofdstukken 3 en 4. In hoofdstuk 3 worden de technische aspecten van de semantiek van vragen en antwoorden kort besproken binnen de theoretische context van de Discourse Semantiek. Exhaustieve interpretatie is geen afzonderlijke operatie toegepast op vragen en antwoorden, maar is geheel geïntegreerd in de gemeenschappelijke dieptestructuur van vragen en antwoorden. In deze gemeenschappelijke dieptestructuur komt het abstracte predicaat besp voor, dat de volgende satisfactiecondities heeft: [[besp]]={<a,b>| a is een argument voor een functie ƒ, b is de bijbehorende waarde}. Met andere woorden, besp vereist een functie ƒ die een argument en een waarde aan elkaar verbindt, en dit is cruciaal: een functie relateert per definitie ieder argument uit het domein aan precies één waarde uit het co-domein. Als gevolg hiervan wordt de uniciteit of exhaustiviteit van een bepaalde gespecificeerde waarde, een bepaal gespecificeerd element uit het co-domein, inherent geasserteerd. De specificatie van een waarde (het beantwoorden van een vraag) is inherent exhaustief.

Maar zoals in hoofdstuk 4 wordt beargumenteerd, ondanks deze integratie van exhaustiviteit in de semantiek van vragen en antwoorden verschijnt exhaustieve interpretatie aan de oppervlakte als een default proces. Ten eerste lijken er pragmatische processen te zijn die de exhaustiviteit beperken; deze pragmatische processen worden in deze studie niet nader uitgezet. Maar ten tweede wordt aan de hand van een aantal voorbeelden betoogd dat er verschillende strategieën zijn die een spreker kan toepassen om aan te geven dat zijn (expliciete) vraag of antwoord als onvolledig, twijfelachtig, of gewoonweg niet exhaustiv moet worden geïnterpreteerd. Deze strategieën kunnen op verschillende manieren gerealiseerd worden, in ieder geval lexicaal of door het gebruik van intonatie. Voor een volledig verslag van de middelen waarover een spreker beschikt om exhaustiviteit te ontkennen of vermijden is echter veel uitgebreider linguïstisch onderzoek nodig; de voorbeelden en overwegingen die in deze studie zijn gepresenteerd kunnen slechts dienen als een eerste indicatie.

De evidentie die bestaat voor de hierboven kort uiteengezette theoretische claims wordt gepresenteerd eveneens in de hoofdstukken 3 en 4. In hoofdstuk 3 worden de getalnamen in detail besproken; echter, deze afzonderlijke bespreking van de getalnamen mag niet worden opgevat als een claim met betrekking tot de specifieke aard van getalnamen. Dit is alleen gedaan om de uiteenzetting zo helder mogelijk te houden. In de hoofdstukken 3 en 4 wordt in het algemeen voor alle semantisch onafhankelijke (mutually exclusive) predicaten, waar ook de getalnamen toe behoren, vastgesteld dat de relatie tussen interpretatie en informatiestructuur als volgt is: wanneer een dergelijk predicaat zich in comment positie bevindt (dus wanneer het in strikte zin het antwoord op een vraag is) krijgt het de "precies" of twee-zijdige lezing. In onze terminologie: het predicaat wordt exhaustief geïnterpreteerd. Dit wordt ondersteund door de observatie dat in een dergelijke constellatie het niet mogelijk is de exclusion inference te negeren of corrigeren zonder dat een contradictie ontstaat. Met andere woorden, de voor implicaturen zo karakteristieke cancellation is onmogelijk, leidt tot een onacceptabele discourse, wanneer semantisch onafhankelijke predicaten zich in comment positie bevinden. Men zou het ook zo kunnen formuleren dat de exclusion inference waarheidsconditioneel is (in tegenstelling tot wat Grice en zijn navolgers beweren). Anderzijds, wanneer een dergelijk predicaat zich in non-comment positie bevindt is een exhaustieve interpretatie niet nodig en is correctie heel wel mogelijk (al brengt dit noodzakelijk een topic-shift met zich mee).

Voor de klassieke voorbeelden van scalaire predicaten (mooi, lelijk, mogelijk, sommige) levert het onderzoek in hoofdstuk 4 andere resultaten op. Exhaustieve interpretatie van een
scalair predicaat zoals mooi heeft niet tot gevolg dat de exclusion inference (in dit geval bijvoorbeeld "niet prachtig") niet gecorrigeerd of genegeerd kan worden. Een dergelijke correctie heeft echter wel altijd een rhetorisch karakter. Correctie van scalaire predicaties in non-comment positie is altijd mogelijk. De voorbeelden die in paragrafen 4.2.1.2 en 4.2.2.2 worden gepresenteerd ondersteunen de conclusie dat de scalaire predicaten semantisch gezien verschillen van de semantisch onafhankelijke predicaten.

De aanwijzingen voor verder onderzoek die naar voren komen uit deze studie hebben voornamelijk betrekking op puur linguistisch onderzoek. Als men het principe dat Horn aanneemt accepteert, en dus ervan uitgaat dat semantische redundantie zoveel mogelijk vermeden wordt in taal, en als men daarnaast de effecten van de informatiestructuur van een discourse op de interpretatie van individuele predicaten zoals hier gepresenteerd accepteert, dan zijn twee typen linguistische gevolgen denkbaar. Het eerste type is distributionele effecten op de volgende manier: een expressie die in isolatie reeds een exhaustieve betekenis heeft (zoals "precies 3" en expressies met alleen (only)) zou niet voorkomen in comment positie, aangezien dan het exhaustieve aspect verdubbeld zou worden. Het tweede type linguistisch effect lijkt echter meer hout te snijden: de verschillen in interpretatie die zijn gevonden tussen predicaten in comment en in non-comment positie gelden voor veel meer predicaten dan die hier zijn besproken. Een interessant kwestie zou zijn de interpretatie van alleen, dat reeds een exhaustief aspect bezit, in comment en in non-comment positie. Opgaven zoals deze staan open voor verder onderzoek.
curriculum vitae