The time course of processing of natural language quantification.

Kevin Brisbane Paterson

Thesis submitted for the degree of Doctor of Philosophy

Department of Psychology
University of Glasgow

December, 1996

Abstract

This thesis examines the processing of quantificational semantics during reading. It is motivated by some formal observations made by Barwise and Cooper (1981), and a psychological theory proposed by Moxey and Sanford (1987, 1993a). Barwise and Cooper classify quantifiers in terms of the directional scalar inference they license. Quantifiers like *a few* and *many* are described as monotone-increasing which means that what is true about a subset is also true of the superset. For instance, if *a few student passed the exam with ease,* this entails that *a few students passed the exam.* Other quantifiers like *few* and *not many* are monotone-decreasing, and license inferences in the opposite direction. If it is true that *few students passed the exam,* this entails that *few students passed the exam with ease.*

Moxey and Sanford found that these categories of quantifier produce contrasting patterns of focus. They used an off-line production task to demonstrate the monotone-increasing quantifiers, like *a few* and *many* focus processing attention on that subset of the quantified NP which is true of the sentence predicate (called the 'refset'), and that subsequent pronouns are interpreted as referring to this set. This means that given a fragment like (1), the plural pronoun will be interpreted as referring to the set of students who passed the exam.

(1) A few of the students passed the exam. They . . .

In contrast, monotone-decreasing quantifiers like *few* and *not many* exhibit a more diffuse pattern of focus, and permit subsequent reference to either the refset, or the complement of this set (called the compset), which is false of the sentence predicate. This means that the plural pronoun in (2) can be interpreted as either referring to the set of student who passed the exam, or those who failed it.

(2) A few of the students passed the exam. They . . .

However, the off-line nature of the Moxey and Sanford studies limit them as descriptions of reading processes, so this thesis reports a series

of experimental investigations of pronominal reference during reading. The first two studies used materials like (3) in a self-paced reading experiment to demonstrate that reference is easier when the anaphor describes a property of the refset (*their presence*) following monotone-increasing quantification, but that reference to either a property of the refset (*their presence*) or compset (*their absence*) is possible following monotone-decreasing quantification, although there is a preference for compset reference.

(3) [A few | Few] of the MPs attended the meeting.
Their [presence | absence] helped the meeting run more smoothly.

A second two experiments monitored subjects' eye movements as the read passages like (3) in order to determine the locus and time course of referential processes. In line with other studies (eg. Garrod, Freudenthal and Boyle, 1993), it was predicted that the anaphor would be immediately interpreted as anomalous when it describes the unfocused antecedent. However the studies failed to find any evidence of punctuate effects. A further eye movement study was conducted with a revised set of materials but still failed to find evidence of punctuate anomaly detection in the two quantificational conditions. It was concluded that pronominal reference to a quantified noun-phrase is not processed on-line, i.e. it is not processed as the anaphor is read.

Chapter eight presents two experiments on the interpretation of the non-monotonic quantifier *only a few*. It was suggested that this has the simple function of marking a set relative to expectations, and that focus is pragmatically determined. Focus is maintained on the refset when the quantified sentence describes a situation which is consistent with expectations, but the compset is placed in focus when these expectations are violated. Experiment six uses a sentence-continuation task to demonstrate these preferences, and an interaction with sentence connectives. Experiment seven monitored subjects' eye movements are they read sentences which referred to either the refset or compset of a sentence quantified by *only a few*. There was no evidence that the processing of pronominal reference is contingent on the focusing properties of this quantifier.

Chapters nine and ten make a digression to consider the interpretation of sentences with more than one quantifier. The resulting scope ambiguity has been the subject of considerable theoretical interest, but limited empirical research. The existing literature is reviewed in Chapter nine, and a preliminary off-line sentence-continuation study is reported in Chapter ten which examines the interaction of quantifier and pragmatic constraints on a doubly-quantified sentence. The experimental findings are summarised in Chapter eleven, where an effort is made to accommodate the quantifier focus and scope ambiguity strands of this thesis within a common representational framework.

Acknowledgements

I wish to thank Dr. Linda Moxey and Prof. Tony Sanford for supervising this thesis. They have been constantly encouraging and supportive.

I thank Dr. Martin Pickering , Dr. Simon Liversedge and Dr. Matt Traxler for their feedback and encouragement. I must also thank Keith Edwards, who maintains the hardware and software in the Glasgow eye movement laboratory, and Eugene Dawydiak for his help in conducting Experiment five.

I also thank Jo, my parents, and those friends in Glasgow who kept me smiling.

I thank the post-graduate students and staff at the Universities of Glasgow and Nottingham for their support; and Prof. Geoffrey Underwood for his patience.

Finally, I must thank the Carnegie Trust for the Universities of Scotland who funded this research through a scholarship, and the ESRC Human Communication Research Centre for financial support in running the eye-tracking experiments.

For my mother and father

Contents

CHAPTER FIVE
The time course of plural anaphoric reference to a quantified noun-phrase.

CHAPTER SEVEN

The locus of referential anomaly detection.

CHAPTER EIGHT
Pragmatic constraints on focus following quantification by only a few

CHAPTER NINE

Constraints on quantifier scope ambiguity

CHAPTER TEN

Experiment Eight:

Semantic and pragmatic constraints on

the interpretation of quantifier scope ambiguity.

CHAPTER ELEVEN
Summary and conclusions

List of tables

List of figures

Chapter one

Chapter two

Chapter three

Chapter four

Chapter five

Chapter seven

Chapter eight

Chapter nine

Chapter ten

Chapter eleven

Chapter one:

Processing consequences of a model-theoretic approach to natural language quantification

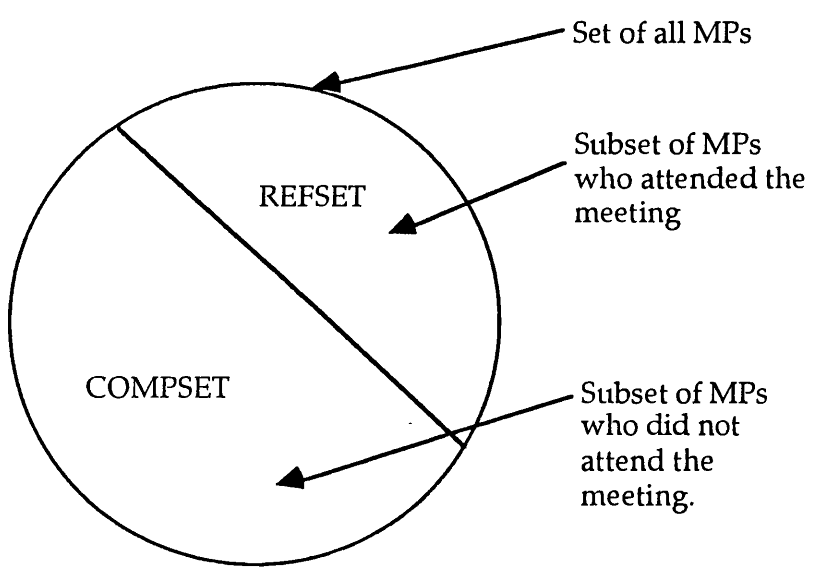
Introduction

The experiments reported in later chapters investigate the interpretation of natural language quantifiers during text comprehension, and take as their starting point a theory of quantifier function proposed by Moxey and Sanford (1987, 1993a). This theory is partly motivated by observations from model-theoretic semantics, and particularly the work of researchers applying developments in Generalised Quantifier Theory (Mostowski, 1957) to linguistics (cf. Barwise and Cooper, 1981; Keenan and Stavi, 1986; Westerståhl, 1989). However, while Moxey and Sanford note some parallels between model-theoretic and psychological approaches to semantics, others have argued for a closer correspondence (cf. Johnson-Laird, 1982, 1983). This chapter assesses the contribution of model-theoretic semantics and generalised quantifier theory to a psychological account of quantificational semantics.

The chapter has two main sections. The first section provides an overview of the model-theoretic approach to natural language semantics, and particularly the approach to natural language quantification. It then concentrates on some attempts to produce a detailed categorisation of natural language quantifiers in terms of semantic properties. This includes the properties of conservativity and extensionality, which define the context-dependency or independency of quantifiers, and the properties of monotonicity and persistence, which are used to describe directional inferences associated with those quantifiers. The most important of these is the property of monotonicity, which is shown to be related to a syntactic phenomena called negativity (Klima, 1964).

The second section of the chapter describes three processing accounts of natural language quantification. The first of these proposes that the processing complexity of quantifiers is explained by the properties of monotonicity and persistence. The second account proposes that processing complexity can be described in terms of the complexity of the automaton that would be required to determine the truth of a quantified sentence. The third account is a psychological account, but is consistent with a categorisation of quantifiers in term of monotonicity. The next chapter will consider the correlation between the semantic property of monotonicity and the psychologically-motivated sub-division of quantifiers described by Moxey and Sanford (1987). In order to set the scene for this comparison, we begin with a brief overview of Moxey and Sanford's account of quantification.

Moxey and Sanford (1987) describe how some quantifiers (eg. *a few* and *few*) are used in partitive constructions of the form *Quantifier of the N VP* to focus processing attention onto contrasting subsets of the quantified noun. They argue that the noun is divided into two subsets, called the **refset** and **compset**; where the refset is that part of the quantified NP of which the VP predicate is true, while the compset is that part of the quantified NP which is false of the predicate. One sub-category of quantifiers (including *a few*) is thought to exclusively focus processing attention on the refset, and another sub-category (including *few*) preferentially focuses on the compset. Figure 1 illustrates the partition of a noun-phrase by *a few*.



Given the sentence *A few of the MPs went to the meeting*, the refset is the subset of MPs who did attend the meeting, and the compset is the complement subset who did not attend the meeting.

Figure 1:
Refset and compset partitions of a plural noun-phrase.

Some experimental evidence reviewed in Chapter 2 suggests that quantifier focus guides coherence processes like anaphoric reference, and can influence the content of inferences. For example, Moxey and Sanford (1987) report sentence-continuation studies where subjects read sentence fragments like (1) or (2), and interpreted the plural pronoun as referring to the focal subset of the quantified NP. This meant that the pronoun was interpreted as reference to the refset of MPs who did not attend the meeting following quantification by *a few*, but tended to be interpreted as reference to the compset following quantification by *few*.

(1) A few of the MPs went to the meeting.
(2) Few MPs went to the meeting,

The content of the sentence continuations also appeared to be determined by the form of quantification. Quantification by *a few* tended to result in sentences which simply continued the narrative from the perspective of the refset, but quantification by *few* resulted in reasons for, or consequences of, the small refset.

Unfortunately, because Moxey and Sanford tested language production, their experimental studies cannot be interpreted in terms of comprehension processes. It may be that the observed focus effects are a part of language production processes involved in generating sentence continuations. Or more plausibly, they may have been due to strategic processing performed in response to the specific task demands, and not representative of normal language processing. The present experiments address questions about the comprehension of anaphora in quantificational contexts, and use methodologies (e.g. self-paced reading and measures of eye movements during reading) which do not draw conscious attention to the experimental manipulation, and are not so prone to strategic processing.

Model-theoretic semantics

The above account, and particularly the model illustrated in Figure 1, depends on the interpretation of NPs, VPs and quantifiers as descriptions of set representations. A similar approach is formalised in model-theoretic semantics (Montague, 1974; see Dowty, Wall and Peters, 1981), where linguistic expressions are translated into set descriptions. This is a complex and non-intuitive account, but one which is fundamental to the semantic properties described later in this chapter, and deserves a detailed introduction.

According to the theory, two intermediary sets of functions called **intensions** and **extensions** are required to translate linguistic expressions into set descriptions. These are the modern equivalents of Frege's (1892) distinction between the sense and reference of an expression. The reference (or extension) describes the objects picked out by the expression in a given state of affairs; while the sense (intension) of an expression is the criteria used to make this selection, whatever the state of affairs. Linguistic expressions are first mapped onto their intension using a general interpreting function (see Figure 2). After this, intensions perform a more complex function of mapping possible worlds (all the conceivable states of affairs) onto the appropriate extension for each of those possible worlds. Finally, extensions pick out the set description of a linguistic expression in a given possible world.

The power of this approach can be demonstrated by a simple example. Consider the expression *the Prime Minister*. Five years ago this would have had the person Margaret Thatcher as its extension. The same expression evaluated today has as its extension another individual with the name John Major. Between these two times the intension of the expression has not changed, the individual who is *the Prime Minister* has been selected using the same criteria, but with different extensional results. Moreover, the two points in time used to evaluate the expression are but two possible worlds, and there are clearly others that could have been used to produce different extensions. For instance, *the Prime Minister* would refer to James Callaghan or Edward Heath at different periods of the 1970's, and it is even possible to think of an imaginary situation where Mickey Mouse is the appropriate extension. In fact, there are an infinite number of possible worlds which can be used by intensions to generate extensions.

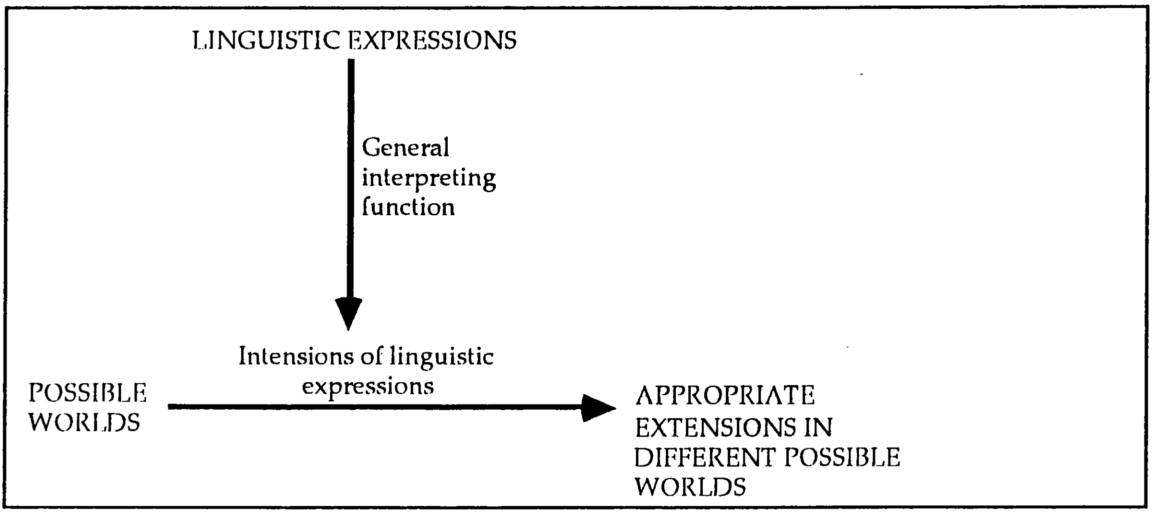


Figure 2: The interpretation of a linguistic expression in model-theoretic semantics. From Allwood, Andersson and Dahl (1977)

While a straightforward extensional semantics is sufficient for most purposes, adding intensions and possible worlds avoids some pitfall originally observed by Frege (1892). The worst of these is the inability of extensional semantics to distinguish between expressions which pick out the same set, even though these expressions are intuitively different. A classic demonstration uses the expressions *morning star* and *evening star*. Although these are both descriptions of the planet Venus, one describes its presence in the morning, and the other describes its presence at night. Because they share the same extension they are treated as semantically equivalent. However, the difference between the two can be captured using intensions and possible worlds. These allow for the existence of an alternative world where the expressions do not share the same extension, but actually describe two separate stars: one that appears in the morning and another which appears at dusk. Since intensions provide the criteria for determining the extension of an expression for each possible world the intensions of *morning star* and *evening star* necessarily differ.

A similar problem occurs when different expressions have the empty set as their extension. Examples of these would be the set of female Presidents of the United States, or dogs who can play Beethoven (Allwood, Andersson and Dahl, 1977). Again, without possible worlds and intensions these descriptions are semantically indistinguishable. With possible worlds the expressions have conceivable non-empty extensions in some alternative world, and therefore must have different intensions.

The introduction of two further principles allow the model-theoretic approach to provide intensions and extensions for complete sentences. The first of these is the principle of compositionality (following Frege, 1892), which asserts that the meaning of a sentence is a function of the meanings of its parts. Montague argues that a functional composition of the intensions of the sentence parts will determine the sentence intension, which can be evaluated against a model to return a truth-value. To achieve compositionality, Montague introduces isomorphic syntax and semantics as a second principle. This is achieved by specifying semantic types which correspond to syntactic categories, and semantic composition rules which parallel the rules of grammar. Johnson-Laird (1983) neatly describes the relationship as semantic ingredients mixed according to a syntactic recipe.

The semantic type assigned to an expression also defines its set description in a given model. For example, the Montagovian equivalent of intransitive verb describes, in a particular model, the set of individuals for whom the verb predicate is true. So a verb like *smiles* will pick out that set of individuals for whom it is true. Similarly, transitive verbs describe a set of ordered pairs, where the predicate is true of each pair. The verb-phrase *is married to* will pick out a set of paired individuals in the model. The treatment of noun-phrases, like *the boys,* is less intuitive. They describe a set of sets, where each of the subsets is the extension of one of the properties of the noun-phrase. So the extension of *the boys* is the set of extensions for all of the predicates which are properties of the noun-phrase. These might include the properties *are male, are young* and *play football.* Proper names like *Tom* also pick out a set of sets, but can be interpreted as the intersection of the set of predicates which form the properties of the proper noun. This will be a single individual in the model (unless it contains more than one person called Tom).

The composition rule for noun-phrases and verb-phrases states that the resulting sentence is true if the set described by the extension of the verb-phrase is a subset of the set described by the extension of the noun-phrase. In other words, the sentence is true if the verb-phrase describes one of the properties of the noun-phrase.

Formal and psychological approaches to semantics

Formal and psychological theories of semantics have different objectives. A formal approach like model-theoretic semantics is concerned with the means of determining the truth of a sentence with respect to a fixed and known model of the world; or to put it another way, it is concerned with specifying the necessary conditions for the truth of a sentence. A psychological theory, in contrast, must explain what model of the world is consistent with the meaning of a sentence, and explain the process of arriving at this interpretation. It must also explain how this interpretative process interacts with other factors. For instance, it must describe how the interpretation assigned to a sentence is further constrained by knowledge about the prior context, the plausibility of alternative interpretations, and the pragmatics of the described situation.

The relationship which holds between formal and psychological accounts may be similar to that which holds between theories of syntax and parsing. Syntactic theory attempts to specify the rules which govern the structure of grammatical sentences, while a theory of parsing describes the computation of that sentence structure. Some theories of parsing (e.g. Frazier and Fodor, 1978) adopt a particular syntactic theory as the description of the output structures, then propose a set of procedures to compute these structures. The parsing theory must be able to explain why one syntactic structure is preferred over another when the sentence is ambiguous, and account for the difficulty experienced when reading particular sentences. This means that a parsing theory will be correct to the extent that the adopted syntactic theory describes the mental representation of syntax, and that the parser is a description of sentence processing.

Some researchers (e.g. Hall-Partee, 1979; Johnson-Laird, 1982, 1983) have argued for a similar approach to semantics. According to this approach, formal semantics is a theory of linguistic competence, which provides an idealised account of the relationship between sentences and models of the world, whereas a psychological theory is an account of linguistic performance. It must describe the extent to which the idealised account can be accommodated within the limitations of the sentence processing system.

Both Hall-Partee and Johnson-Laird acknowledge that the impossibility of cramming the infinite number of possible worlds into the finite capacity of a brain is a major performance restriction. They propose two ways around this problem: either there is a limit on the number of possible worlds that can be mentally represented, or there is a set of procedures to generate possible worlds as they are needed. Hall-Partee and Johnson-Laird favour the latter solution. This introduces a secondary problem of whether a language user with limited access to possible worlds can be aware of the intensions of words. Intensions were defined in the previous section as a function from the set of all possible worlds to extensions in a given world. However, these intensional functions cannot work if there is restricted access to the possible worlds. Hall-Partee suggests that the problem is avoided by allowing partial functions as intensions. This implies a language user who is not unaware, but has an imperfect awareness of the intensions of words, which concords with informal observations by Johnson-Laird (1982). He notes that while a language user might claim to fully understand a particular sentence, she will have greater difficulty in defining many of the component words.

Other performance restrictions are appropriate for the representation of a single model. We have already seen the complexity of the formal semantic treatment of noun-phrases, where they are represented as a set of all the properties which are true of the noun-phrase. It is unlikely that this degree of detail is mentally represented, if only because it requires more than the available partial intensions. Hall-Partee (1979) suggests that only partial mental representations are constructed and that "communication is possible as long as there is sufficient similarity in our partial models and our imperfect semantics" (Hall-Partee (1979, p39).

The degraded nature of mental representations is easy to demonstrate. Johnson-Laird (1982) distinguishes between the logical completeness of a formal semantics model, where there is a determinate truth-value for any sentence evaluated against it, and what he refers to as the 'radical incompleteness' of mental representations. Often the truth of a sentence is indeterminate with respect to a mental model because of the relative poverty of the representation. Johnson-Laird argues that quantified sentences like (4) have an indeterminate truth-value relative to the model described by (3), because (3) can describe a situation where the difference between the set of smokers and set of vegetarians is the empty set, or one where the difference is a set of smokers who are not vegetarian.

(3) All of the vegetarians are smokers.
(4) Some of the smokers are not vegetarians.

Sentence (3), he suggests, is represented either by a number of separate models, or by a simultaneous representation of the sets of possible states or affairs, and a model of the determinate aspects of the discourse. Spatial descriptions offer further evidence of indeterminacy that require multiple or vague representations (cf. Mani and Johnson-Laird, 1982; Johnson-Laird, 1983), as does multiple quantification. For example, a sentence like (5) is ambiguous between at least two interpretations, one where *every man loves the same woman*, or one where *every man loves some or other woman*. Multiply-quantified sentences are further discussed in Chapters nine and ten.

(5) Every man loves a woman.

These examples describe the indeterminacy of some quantified sentences. However, a separate hypothesis will be pursued in this thesis; it will explore the claim that quantifiers are used to direct the construction of mental representations, in order to achieve the shared partial representations which Hall-Partee believes to be necessary for successful communication. Johnson-Laird and Byrne (1991) report an experiment on syllogistic reasoning which illustrates what this means. Their subjects read syllogisms where the first premise was either quantified by *all* (6) or *only* (7). These two sentences have the same truth conditions, but Johnson-Laird and Byrne predicted that the mental representation of (7) included the explicit information that anyone who is not a baker is also not an athlete, whereas this remains indeterminate in the mental representation of (6).

(6) All of the bakers are athletes.
(7) Only the bakers are athletes.

This prediction was supported by the finding that subjects who then read a second premise like (8) were more likely to conclude that Mark was not an athlete when this premise followed (7) than when it followed (6).

(8) Mark is not a baker.

It may be that the mental representations for sentences quantified by *a few* and *few* differ from each other in a similar way. A quantified sentence like (9) might produce a mental representation of the refset of MPs who attended the meeting, but remain indeterminate about the compset of MPs who were absent. This would explain the difficulty in making pronominal reference to the compset. However, the mental representation of a sentence (10) might include both the refset and compset, and so enable reference to either subset.

(9) A few of the MPs went to the meeting.
(10) Few MPs went to the meeting,

The remainder of this chapter will focus on formal approaches to natural language quantification, then consider three attempts to translate these into processing accounts.

Model-theoretic quantification

Montague (1974) extends the model-theoretic approach to include a radical treatment of quantification which preserves the syntactic and semantic uniformity apparent in natural language. The more traditional logic of predicate calculus[1] describes two quantificational operators which express the mathematical properties of existence and universality. These correspond only loosely to the natural language descriptions *at least one* and *all*. The quantifiers prefix logical formulae and are used as a means of introducing variables. For instance, in (11) the universal quantifier ($\forall$) binds the variable 'x' which appears in the formula as the operand of the function 'R', to assert that for all substitutions of the variable 'x', the application of function 'R' is true. Longer formulae might include multiple occurrences of the variable which are all interpreted as bound by the quantifier.

[1] cf. Andersson et al, (1977), McCawley (1981).

(11) $\forall x\ (Rx)$

Translating a quantified sentence into predicate logic disrupts the syntactic form of the original. For example, (12) has (13) as its logical interpretation. Here the quantifier *some* is interpreted (imprecisely) as the existential quantifier ($\exists$), and the noun and verb-predicate order is reversed. The quantifier is placed outside the formula as a means of introducing the variable *boy*.

(12) Some boy runs.

(13) $\exists$ boy (RUN(boy))

In contrast, Montague attends to the structure of natural language to treat quantifier syntactic operators on nouns like *boys* to produce noun-phrases like *some boys*. Semantically, they no longer prefix formulae as a means of introducing and binding variables, but enter into the formulae as functions on nouns. We have already discussed the extensional semantics of verb-phrases and noun-phrases as properties and sets of properties respectively. A combination of an NP and VP produces a sentence, which is true if the property which is the extension of the VP is a member of the set of properties which are the extension of the NP. Quantified NP's have as their extension a superset of the properties that are true of the unquantified noun. For example, it might be true of the plural noun *men* that they *run* and *jump*; but *some man* might only *run* and not *jump*. As with unquantified NPs, the combination of quantified NP and VP is true if the property described by the VP is a member of the set of properties denoted by the quantified NP.

The uniformity of this approach is attractive, but there are some problem cases (Westerståhl, 1984, 1989). The most immediate of these are examples of quantifiers which do not operate on nouns, but instead appear to act anaphorically. Westerståhl argues that sentences like (14), (15) and (16) are instances where the quantifiers take a 'dummy' argument whose content is derived from context.

(14) Some like it hot.

(15) Few come to visit anymore.

(16) All cheered.

Further examples are complex quantifiers like *something* and *everything* which resemble quantifiers but do not fill the same syntactic role unless they are treated as complex expressions which decompose to *some* + *thing* and *every* + *thing*.

Generalised quantifier theory

Barwise and Cooper (1981; also Keenan and Stavi, 1986) use this general model-theoretic framework as the basis of a more detailed classification of quantifiers. Not all of the classifications are relevant to the present issue, and attention will be restricted to those which characterise the function of quantifiers relative to context, and others which describe their inferential properties. At the same time, it must be remembered that while researchers broadly agree about the classification of quantifiers, there are some disputed cases. For instance, Barwise and Cooper differ over judgements about the quantifiers *few* and *many*, which Keenan and Stavi exclude from their analysis. Keenan and Stavi also have problems with *a few*. Westerståhl (1989) proposes different interpretations for *many* in an effort to resolve the difficulties it raises.

These quantifiers are of interest precisely because of the problems they present for semantic theories. Often their interpretation appears to be context-dependent. They are also quantifiers investigated by the present experiments, so are used as examples throughout the discussion of semantic categorisation schemes. Another quantifier, *only a few*, is also included, and is the subject of some experiments reported in Chapter 8.

Conservativity and extensionality

We begin with the related properties of **conservativity** and **extensionality**. Conservativity states that a quantifier functions over a limited domain, and that the truth of a quantified sentence depends only on that part of the VP extension which is common to the noun of the quantified NP (Westerståhl, 1989). This means that no other NP sets of which the VP is true share any responsibility for the truth of the sentence. For example, the truth of a sentence like *every sailor survived* is only dependent on the set of sailors who survived, and no other sets of surviving entities. Extensionality requires that a quantified sentence maintains its truth-value under expansion of the model. This means that the truth of the sentence depends only on those aspects of the model describes by the sentence, and that changes to other aspects of the model will not affect the truth of the sentence. The precise implications of these constraints will become clearer in a moment, but it is sufficient to note that they require that the truth of a quantified sentence is independent of any other aspects of the model, which means that its interpretation is context-independent.

A quantifier is judged to be conservative or extensional from intuitions about its performance in linguistic tests (Barwise and Cooper, 1981). The same method can be used to show that some of the presently considered quantifiers are borderline cases for the conservativity constraint (particularly *only a few*) and most of them are non-extensional. The test sentences are based on a scenario where a ferry carrying a crew of sailors and some passengers has sunk.

If conservativity is a property of the quantifiers used in sentences (17) to (22) then the truth of these sentences must depend only on the set of surviving sailors, and no other sets of survivors. This is clearly the case for (17) and probably (18), but (19) to (22) are more contentious. On one reading, the quantifiers in these examples describe the set of surviving sailors alone, but on another reading they describe a quantified amount of surviving sailors relative to the total number of survivors. This second reading is particularly strong for *only a few* (sentence 22). It seems to assert that either *only a few* sailors survived relative to a much larger number of passengers, or no passengers survived. More will be said about this comparative function when I discuss extensionality.

(17) Every sailor survived.
(18) A few sailors survived.
(19) Few sailors survived.
(20) Many sailors survived.
(21) Not many sailors survived.
(22) Only a few sailors survived.

Westerståhl (1984) argues that the conservativity restriction should be broadened from quantification over the noun set of the sentence to quantification over '**context sets**' in the discourse. These are contextually selected sets which act as extensions for quantified NPs. In most cases the context set equals the noun set (Westerståhl calls this context set the **NP-universe**), but there are exceptions, perhaps including some of the examples discussed above, and certainly including instances of pronominal quantifiers. If we imagine a courtroom scenario where a wrongly-convicted person has just been sentenced, then the following are possible fragments of discourse about the event:

(23) The judge hammered his gavel. All stood up.
(24) A crowd of people jeered. All wanted justice.

In (23) *all* refers to all of the people in the courtroom (excluding the judge), which Westerståhl calls the discourse universe. With (24) it refers only to the crowd of people who jeered, not the full discourse universe.

The reference of quantified NPs can also be contextually constrained. Westerståhl suggests the following example (25), where *most children* must refer to the set of English children, even though the quantified NP implies an NP universe of all children, and the passage has a discourse universe which includes the children from many countries. Finally, he argues that the definite article selects the context set for partitive constructions like *Quantifier of the N.*

(25) The English love to write letters. Most children have several penpals in many countries.

The quantifiers in sentences (17) to (22) have two possible contexts. Either they operate on the NP-universe, the set of sailors, or else they operate on a broader context set containing all of the individuals in the discourse. This set contains both the sailors and the passengers on the ferry.

A similar pattern of results are observed on tests of extensionality. These require that the truth of a sentence is dependent only on those aspects of the model which are explicit in the sentence. A quantifier is demonstrably non-extensional if sentences like (17) to (22) are possibly false, despite the truth of sentences (26) to (31). By this criteria *every* is clearly extensional, as (17) and (26) are synonymous; but the truth of the other quantifiers depends on what the set of surviving sailors is compared against. In sentences (27) to (31) this is clearly the full set of sailors, but the comparison set is not so obvious for sentences (18) to (22). For example, on some readings of (20) there might be *many* surviving sailors relative to the total number of survivors, or relative to the total number of those originally on board the ferry. Neither of these readings are synonymous with (29), and imply the non-extensionality of *many*. Similar alternative readings are possible for *not many, few* and *only a few*. According to my interpretation *a few* is extensional, although Keenan and Stavi (1986) claim that others find it non-extensional.

(26) Every sailor was a surviving sailor.
(27) A few sailors were surviving sailors.
(28) Few sailors were surviving sailors.
(29) Many sailors were surviving sailors.
(30) Not many sailors were surviving sailors.

(31) Only a few sailors were surviving sailors.

Westerståhl (1989) formalises the observation that *many* can have several readings, some of which are non-extensional. He proposes three alternative interpretations. In one *many* compares the set it quantifies (the *many* of the NP under VP predication) to the size of the overall universe. In another, the comparison is a 'normal', or expected, frequency for the noun quantified by it. A third interpretation compares the quantified NP to the overall frequency of the predicate in the universe, although Westerståhl rejects this option because it also implies non-conservativity.

Further evidence of the non-extensionality of *many* is given by Keenan and Stavi (1986). Their argument depends on the reader following a lengthy story about a rather complicated scenario, but has the advantage of not relying on the reader to agree with their intuitions about sentence interpretation. Readers are asked to imagine an annual meeting of some society which is normally attended by a large number (say ten thousand) people who are qualified as doctors, and another one or two who are lawyers. In the year prior to one of these meetings all of the doctors become qualified lawyers, and the lawyers simultaneously become qualified as doctors, with the result that everyone is both a lawyer and a doctor. Now Keenan and Stavi imagine that the overall attendance at this year's meeting is much lower than usual: only about 500 of the doctor/lawyers turn up. Keenan and Stavi suggest that, although the sets picked out by the two expressions *the doctors* and *the lawyers* are semantically equivalent - they have the same extensions - sentence (32) will be true and sentence (33) false. This is because the attendance is judged relative to a norm attendance, and while this year there are many more lawyer than normal, there are also many less doctors. Because of this the quantifier is necessarily non-extensional. Keenan and Stavi argue that a variant of this example will work as well for *few* and, by implication, *not many*. Substitution of *only a few* for *few* suggests that these are similarly non-extensional.

(32) Many lawyers came to the meeting this year.

(33) Many doctors came to the meeting this year.

It was noted earlier that conservativity and extensionality imply context-independency. The converse argument claims that quantifiers which fail to observe these properties will have a context-dependent function. This is the conclusion drawn by Westerståhl (1989), who observes that his two non-extensional meanings of *many* require a comparison against some norm frequency which is not represented in the model. He also notes that the standard of comparison varies across contexts. For example, the sets of boys for (34) and (35) are *many* with respect to different criteria. In (36) the criteria vary within the one sentence. Here the *many boys* are compared against the universe of boys (which is an extensional interpretation), but *many girls* is compared against the expected frequency of girls dated by boys.

(34) Many boys in the class are right-handed.
(35) Lisa is dating many boys in the class.
(36) Many boys date many girls. (Partee, cited in Westerståhl, 1989)

However, Keenan and Stavi contradict the assumption that non-extensionality necessitates context-dependency. They argue that sentences quantified by *many* or *few* have the narrower function of expressing the speaker's belief that an amount is significant. Such sentences have indeterminate truth-values because the reader cannot know the standard of comparison intended by the writer. They use sentence (37) as an example.

(37) Many tourists visited the zoo today.

If we imagine that (37) is uttered in the context of a rainy bank holiday during the summer, then Keenan and Stavi observe that the sentence might be true if the number of tourists is compared against the number who would normally visit the zoo on a rainy day, but plausibly false if compared against the number who would normally visit on a summer bank holiday with better weather. Because the listener cannot establish which of these comparisons is intended, she cannot know the truth of the sentence, only something about the speaker's beliefs. From a more psychological perspective this objection is less problematic. I earlier agreed with Hall-Partee (1979) that the purpose of language comprehension is to construct partial representations of a model by interpreting the descriptions given by the writer, and that the closer this approximates the model held in the writer's mind, then the more effective the communication. Successful communication will mean the intended standard of comparison of quantifiers like *many* and *few* is easily recovered from the context.

Monotonicity and persistence

The second two properties of monotonicity and persistence[2] label patterns of inference associated with a quantifier within a single model. Barwise and Cooper (1981) highlight some processing implications of these two properties, while Moxey and Sanford (1987, 1993a) suggest that the focus effects they describe are processing correlates of monotonicity.

Monotonicity describes the direction of inferences about the truth of quantified sentences under the expansion or contraction of the VP predicate. Quantifiers which are monotone-increasing permit scalar inferences about the truth of a superset description given the truth of a subset description. Those which are monotone-decreasing permit inferences in the opposite direction. A third category of quantifiers are non-monotonic and fail to preserve the truth of inferences in either direction. Persistence is an equivalent description of inference patterns for quantified NPs under expansion or contraction of the noun-phrase set under constant VP predication. Quantifiers are either persistent and permit scalar inferences from NP subset to NP superset, anti-persistent and permit inferences in the opposite direction, or non-persistent and block both sets of directional inferences.

I will begin with a classification of the monotone properties of quantifiers. This depends on intuitive judgements about the truth of sentences like (38) to (41). These illustrate tests of upwards monotonicity (38 and 39) and downwards monotonicity (40 and 41) for the quantifiers *some* and *no.* The acceptability of (38) and unacceptability of (39) demonstrate the classic monotone-increasing property of *some,* while the reversed pattern of acceptability for (40) and (41) describe *no*'s downwards monotonicity. The directional inferences entailed by these tests are illustrated in Figures 3 and 4.

(38) If *some* student passed the exam with ease, then *some* student passed the exam.

(39) × If *some* students passed the exam, then *some* students passed the exam with ease.

(40) × If *no* students passed the exam with ease, then *no* students passed the exam.

(41) If *no* students passed the exam, then *no* students passed the exam with ease.

[2] Monotonicity and persistance are also known as left and right monotonicity (cf. van Eijk, 1986) which captures the closeness of the two properties but can be confusing. We retain Barwise and Cooper's terminology for the sake of some clarity.

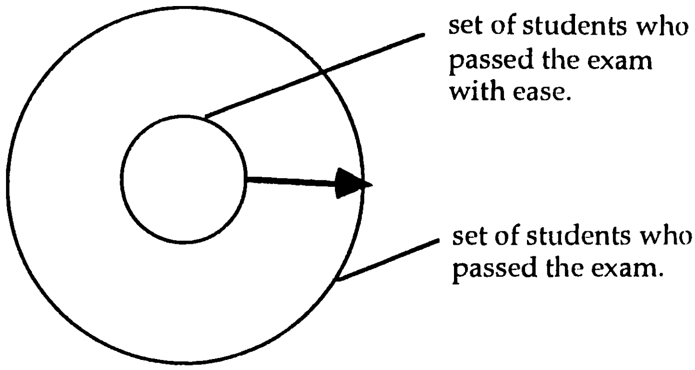


Figure 3: Illustration of monotone-increasing inference.
A quantifier is monotone-increasing if it permits inferences about the truth of the VP superset given the truth of the VP subset.

Substitution of the quantifiers *a few* and *many* produce acceptable sentences on tests of upwards monotonicity. In contrast, substitution of *few* and *not many* result in acceptable sentences for tests of downwards monotonicity. My intuitions about *few* and *many* are supported by observations by Barwise and Cooper (1981) and Moxey and Sanford (1987, 1993a), while Moxey and Sanford agree that *a few* is monotone-increasing. Barwise and Cooper argue that *a few* can be both monotone and non-monotonic, depending on its interpretation. If it is understood to mean *some but not many* then *a few* is non-monotonic, but if it means *at least a few* then it is monotone-increasing.

Judgements over the monotonicity of *only a few* are also divided (see 42 and 43). Moxey and Sanford describe *only a few* as monotone-decreasing, but my intuitions are that *only a few* is non-monotonic and does not produce acceptable sentences on tests of either upwards or downwards monotonicity. Other non-monotonic quantifiers are *exactly two* and *exactly half.*

(42) × If *only a few* students passed the exam with ease, then *only a few* students passed the exam.

(43) × If *only a few* students passed the exam, then *only a few* students passed the exam with ease.

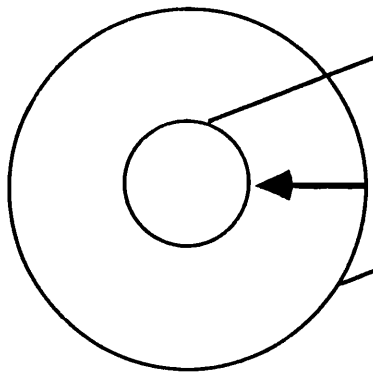


Figure 4: Illustration of monotone-decreasing inference.
A quantifier is monotone-increasing if it permits inferences about the truth of the VP subset given the truth of the VP superset.

The same sort of tests are used to categorise quantifiers in terms of their persistence. Sentences (44) and (46) are tests of persistence for the quantifiers *some* and *no* respectively, and sentences (45) and (47) are the corresponding tests of anti-persistence. The acceptability of (44) and unacceptability of (45) demonstrates the persistence of *some.* The anti-persistence of *no* is clear from the unacceptability of (46) and acceptability of (47). As before, the inferences evaluated by these sentence pairs are illustrated in Figures 5 and 6.

(44) If *some* short students passed the exam, then *some* students passed the exam.

(45) × If *some* students passed the exam, then *some* short students passed the exam.

(46) × If *no* short students passed the exam, then *no* students passed the exam.

(47) If *no* students passed the exam, then *no* short students passed the exam.

Substituting the quantifiers *few,* and *many* into these tests does not yield the same clear results as tests of monotonicity, while the behaviour of *a few* and *only a few* is not considered by other researchers. Barwise and Cooper (1981) tentatively class *few* as anti-persistent and, with even more doubts, class *many* as persistent. My own intuitions are that *few* is clearly not persistent but

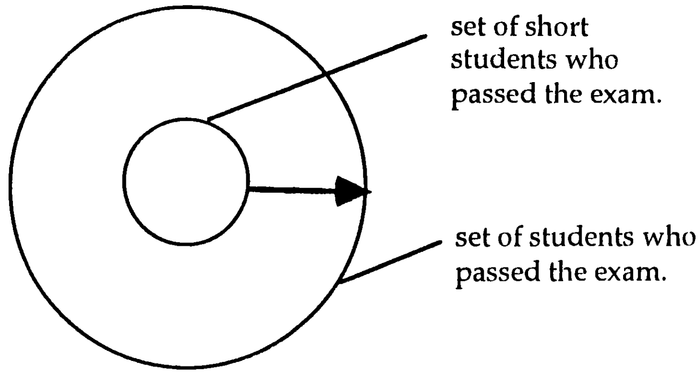


Figure 5: Illustration of persistence inference.
A quantifier is persistent if it permits inferences about the truth of an NP superset given the truth of the NP subset.

difficult to classify under tests of anti-persistence unless it is understood to mean *x or less* including zero. For *many*, it depends on the standard of comparison. If *many* is understood extensionally then it is persistent, but if a non-extensional interpretation is used then it fails both tests. *Not many* seems to behave similarly to *few*, and again requires an *x or less* interpretation to be classed as anti-persistent. *A few* appears persistent, and *only a few* is again difficult to classify. It fails tests of persistence (48) but is questionable on tests of anti-persistence (49). Moxey and Sanford suggest that *only a few* is best interpreted as *x and no more*, which is both non-persistent and non-monotonic.

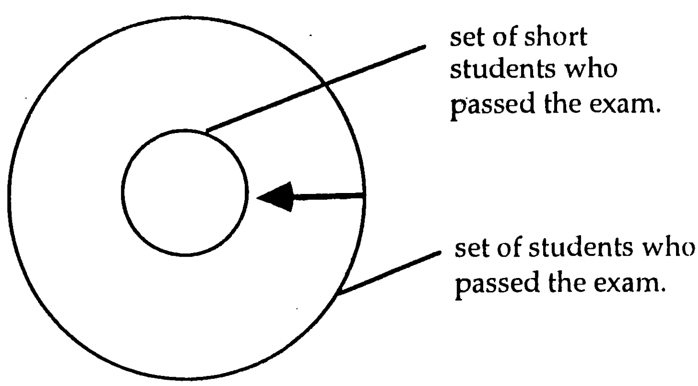


Figure 6: Illustration of anti-persistence inference.
A quantifier is persistent if it permits inferences about the truth of an NP subset given the truth of the NP superset.

(48) × If *only a few* short students passed the exam, then *only a few* students passed the exam.

(49) ??If *only a few* students passed the exam, then *only a few* short students passed the exam.

Monotonicity and negativity

The observations of monotone inference patterns derive from a more general attempt to characterise negativity (cf. Klima, 1964; Fauconnier, 1978; Ladusaw, 1979). Both Fauconnier and J. D. Fodor (reported in Ladusaw, 1979) discovered that constructions previously labelled as negative or affective have the ability to reverse directional scales of implicature and inference associated with the positive versions of those constructions.

The starting point for these observations was given by Klima (1964), who provided a set of formal tools for categorising lexical items which are commonly recognised as negative, but do not have an explicit negative component. He noted, for example, that native speakers of English are aware that *never* is equivalent to the explicitly negative *not ever*, and that *fail* is the negative pole of the lexical pair *fail* and *succeed*. Other examples of positive and negative components are found across a range of syntactic classes including quantifiers, quantificational adverbs, prepositions, adverbial constructions, verbs and adjectives. According to Klima, the negativity of an item is demonstrable by substituting them into declarative sentences with tag questions or negative polarity items (see Zwarts, to appear, for a treatment of negative polarity items).

Klima uses tags of the form " . . . don't they?" and " . . . do they?" to distinguish positive and negative declarative sentences respectively, but without any clear theoretical motivation. He observes that those sentences which are intuitively positive (e.g. 50) accept the negative tag "don't they" and not the positive tag "do they"; and the converse is true for sentences which are explicitly marked as negative (e.g. 51).

(50) All of the students study logic, don't they?

(51) None of the students study logic, do they?

The same sentence frames can be used to test the polarity of other quantifiers. For instance, monotone-increasing quantifiers like *a few* and *many* produce acceptable sentences when the tag question is a test of positive polarity (e.g. 52), and unacceptable sentences when it is a test of negation (e.g. 53).

(52) A few / Many of the students study logic, don't they?

(53)* A few / Many of the students study logic, do they?

Conversely, the monotone-decreasing quantifiers *few* and *not many* produce acceptable sentences when the tag question is a test of negativity (e.g. 54) and unacceptable sentences when it is a test of positive polarity (e.g. 55).

(54)* Few / Not many of the students study logic, don't they?
(55) Few / Not many of the students study logic, do they?

It should be noted that these judgements are not so easy to make as for test frames containing the logical quantifiers *all* or *none*, but the intuitions reported here are in agreement with both Klima, and Moxey and Sanford (1993a). It is even more difficult to make judgements about the polarity of *only a few* using tag questions. It appears most acceptable within sentence frames where the tag question is a test of negative polarity (56), and of borderline acceptability on tests of negativity (57). Moxey and Sanford (1987) argue that sentences which test for positive polarity appear odd when they contain the quantifier *only a few*; but that the corresponding negative test frame is entirely unacceptable.

(56) Only a few of the students study logic, don't they?
(57)? Only a few of the students study logic, do they?

Negativity is also demonstrated by the acceptability of lexical items within declarative items containing negative polarity items like *ever*, *any* or *anymore*. Klima observed that these sentences appear acceptable when they contain negative lexical items and unacceptable when they contain positive lexical items. Ladusaw gives an extreme example of a sentence containing negative polarity items (58) which can be used to test the polarity of quantifiers.

(58) QUANTIFIER of the Chrysler dealers **ever** sell **any** cars **anymore**.

The substitution of an explicitly negative quantifier like *none* is produces an acceptable sentence, as does *few* and *not many*. However, neither *all, a few* or *many* produce acceptable sentences, while *only a few* appears to be a borderline case.

Fodor (also Ladusaw) observed that those quantifiers classed as negative by tag tests and their compatibility with negative polarity items exhibit the downwards direction of entailment which Barwise and Cooper call downwards-monotonicity. Fauconnier argues that the semantic entailments associated with negative lexical items are but one example of scalar expressions. He states: "negation plays no special role . . . it is simply one of the many scale-reversing environments (perhaps a statistically dominant one)" (Fauconnier, 1978, pp295).

Fauconnier's own examples of scale reversal are pragmatic. He argues that grammatical superlatives of the form **the most Adj N** or **the Adj+est N** licence directional pragmatic implicatures. For example, (59) licences the implicature that Max can solve problems ranging downwards across mediumly difficult problems to the easiest one, which is captured by (60). Sentence (61) permits implicatures in the opposite direction: if Max cannot solve the easiest problem, then neither can he solve mediumly difficult or hard ones, or any problem (e.g. 62). Because these scalar patterns are pragmatic in origin they are defeasible, meaning that their implicatures are cancellable, unlike the semantic entailments describe by Fodor.

(59) Max can solve the most difficult problem.
(60) Max can solve any problem.
(61) Max cannot solve the simplest problem.
(62) Max cannot solve any problem.

Moxey and Sanford (1993a) observe that positive and negative quantifiers can be ordered in pragmatic scales which resemble those described by Fauconnier, and believe that these scales reflect the strength of claim associated with a quantifier. For example, they use (63) and (64) to demonstrate a negative scale running downwards from *not many* through *hardly any* to *none*. This means that *not many* implies that *hardly any* may be the case, and both imply that *none* may be the case, but *hardly any* is a closer approximation to *none* than *not many*. This is clear in the unacceptability of the implicature described in (66).

(63) Hardly any of the students study logic, in fact none do.
(64) Not many of the students study logic, in fact hardly any do.
(66)* Hardly any of the students study logic, in fact not many do.

This can be re-stated in terms of focus. *None* is used to focus on the compset, and requires that the refset takes an empty set interpretation. The other quantifiers, *hardly any* and *not many* also focus on the compset but do not require that the refset is empty. Instead they imply that this may be the case, but differ in the strength of this claim. *Hardly any* strongly asserts the empty refset compared to *not many*.

The directional inferences observed by Fodor, and the directional implicatures described by Fauconnier are both examples of perspective control, which contribute to the maintenance of shared representations between language-users. Moxey and Sanford show how these patterns of inference and implicature can be interpreted in terms of focus for quantifiers, while an empirical study by Jarvella and Lundquist (1994; described in the next chapter) demonstrates focusing effects for pragmatic scales. Other syntactic classes which have negative members may also exhibit focusing properties. This is not entirely hypothetical since Moxey, Sanford and Barton (1990) report focus effects for quantificational adverbs.

The processing complexity of quantifiers

In the previous section, a distinction was drawn between the comprehension of a sentence, and the verification of its truth. It was argued that only the former is an automatic part of the reading process. However, most research into natural language quantification has concentrated on truth verification, and asked questions such as whether the semantic complexity of a sentence correlates with the time taken to determine its truth. This section describes three accounts of the processing of sentence truth. A final section considers the relevance of these accounts to a theory of comprehension.

Witness sets and monotonicity

Barwise and Cooper (1981) predict that monotonicity has processing consequences for the verification of sentence truth. They first note that Montague's approach to NP representation and the verification of simple sentences is psychologically unrealistic. According to Montague, an NP is denoted as the family of sets to which that NP belongs, and a simple sentence is true if its VP denotes one of these sets. This means that an NP such as *the men* has as its extension all those sets to which it belongs, including the set of things which are male, and the set of things which are people. If interpreted as a processing theory, it predicts that a reader verifies the truth of a simple sentence like *the men run* by first of all calculating the denotation of *the men*, then determining if the set of runner belongs to this denotation.

Barwise and Cooper propose a more parsimonious account, where a sentence is true if it is possible to find a 'witness set' for which the verb-phrase predicate is true. A witness set is defined as a subset of the noun-phrase denotation such that the noun-phrase is true. For example, a witness set for *some men* is any subset of some men. Similarly, a witness set for *a few men,* is any subset composed of a few men, and a witness set for *few men* is any set of men who are few in number. In any given model there may be may subsets which constitute a witness set, and the truth of a sentence like *some men run* is established by finding a witness set of some men such that those men run. If such as set cannot be found, then the sentence is false.

Witness sets specify the minimal conditions for the truth of a quantified sentence. For instance, it might be the case that all of the men run in the model used to verify that *some men run,* but it is sufficient to identify a witness set composed of some men. This is complicated by the fact that different minimal conditions hold for the truth of monotone-increasing and monotone-decreasing quantified sentences. The minimal conditions for the truth of a sentence like (67) is that no less than a few men run. The sentence is still true in models where more than a few, or even all of the men run. It follows that to evaluate the truth of a monotone-increasing quantified sentence, it is sufficient to find a witness set for which the VP is true.

(67) A few men run.

However, the minimal conditions for the truth of a sentence like (68) is that no more than few men run. To calculate this it is necessary to find a witness set of few men who run, then check that no other men run. If any are found, the sentence is falsified.

(68) Few men run.

If these procedures are psychologically real, then the added complexity of the procedure used to verify the truth of monotone-decreasing quantified sentences implies a greater processing complexity, which will be evident as a longer decision latency in experiments where subjects are asked to verify the truth of quantified sentences.

Barwise and Cooper extend this account by arguing that non-monotonic quantified sentences like (69) require more complex verification procedure, and longer decision latencies, than either monotone-increasing or monotone-decreasing quantified sentences.

(69) Exactly two men run.

This is because non-monotonic quantifiers are a conjunction of monotone ones. For example, *exactly two* is a conjunction of the monotone-increasing *at least two* and the monotone-decreasing *at most two*. According to Barwise and Cooper, this means that two separate procedures are required to verify the truth of a non-monotonic quantified sentence: one to evaluate the monotone-increasing component, and a second procedure to evaluate the monotone-decreasing component.

However, this is unnecessary because the monotone-increasing procedure is included within the monotone-decreasing one, so only the latter is required to verify the truth of a non-monotonic quantified sentence. For example, to verify the truth of (69), the reader must identify a witness set of two men, which satisfies the requirement that at least two men run, then check that no more than two men run.

There are more general problems with the claim that these procedures predict the processing complexity of quantifiers. For instance, the above account predicts that a monotone-increasing quantifier like *all* will have a shorter verification time than a monotone-decreasing quantifier like *no*, but this is demonstrably wrong. To verify the truth of either (70) or (71) requires that the entire set of men is inspected.

(70) All men run.
(71) No men run.

Instead of claiming a direct correlate with processing complexity, it may be sufficient to suggest that the verification of monotone-increasing and monotone-decreasing quantified sentences draws on separate procedures. A monotone-increasing sentence is most efficiently verified by searching for supporting evidence, but the verification of a monotone-decreasing one depends on a search for counter-evidence.

This is most easily demonstrated by considering a more detailed example. Imagine a situation where a carpark attendant is asked to check, as quickly as possible, whether a certain amount of cars in the carpark have their hand-brakes in the off-position. To establish that a monotone-increasing number of quantifiers have their hand-brake off, the most efficient strategy is to successfully identify such cars until the quantifier is satisfied. So if told that there are some cars with their handbrake off, the attendant would search until some such cars were found. However, if told that a monotone-decreasing amount of cars had their hand-brake off, then the attendant should search until either the sentence was falsified or every car has been checked. For example, if told that there were *no* cars with their hand-brake off, the attendant should check all of the cars until one is found with its hand-brake off. Similarly, if asked to check if exactly two cars have their hand-brake off, the attendant must search until two such cars are found, then continue to check that there are no others. We can therefore conclude that the procedure neeed to verify the truth of a monotone-increasing quantified statement is less complex that that needed to verify the truth of a monotone-decreasing quantified statement. However, for the present example at least, there is no difference in complexity for procedures needed to verify the truth of monotone-decreasing and non-monotonic decreasing quantified sentences.

Semantic automata and processing complexity

An alternative account of processing complexity (van Bentham, 1987) considers what procedures are necessary to calculate the truth of a quantified sentence of the form Q A B. Van Bentham categorises quantifiers by the type of abstract mathematical machines (called automata) which could carry out these procedures.

The best-known type of automaton is described by Turing (1944) to express the universality of computation, where any operation which can be expressed as a series of procedures can be carried out by an abstract computing device. This account is a theoretical forerunner of present-day computers and programming languages. A Turing machine reads from a list of symbols (imagine a ticker tape stream of symbols fed into a computer) where the current symbol produces a change in the internal state of the machine. The machine then moves onto the next symbol to produce another change in its internal state. At each point the internal state change is a function of the current state of the machine and the symbol read from the tape.

Van Bentham uses two types of automata to describe the procedures entailed by different quantifiers. Finite-state automata are the simplest form of Turing machine, and can read only the current symbol, change their internal state, and move on to the next symbol. Laport (1994) suggests an elevator control as an exemplar finite state automaton. It keeps record of the current position of the elevator (the machine state) and the nature of the current request. From this knowledge it can choose between two types of possible response, and send the lift either up or down.

The second type of automaton described by van Bentham are push-down automata, and work in the same way as finite-state ones, but with the addition of a stack as an internal memory structure. A stack is a device where responses to the current input can be recorded and stored for later use, or previously recorded responses can be erased, depending on the nature of the current input. Its major constraint is that items on the stack can only be removed from the top, following a last in, last out protocol; but there may be other constraints on the readable depth of the stack. For instance, the automaton may only be able to read the top item, or perhaps the top two items on the stack, at any point in time.

Both types of automata operate in the same way as a Turing machine. They are presented with a string of symbols which describe whether a member of the noun set A in a given model is also a member of B. Membership of B is symbolised by '1', and non-membership is symbolised by '0'. The automaton for each quantifier reads this string of 1's and 0's, alters its internal state accordingly. The final state of the automaton represents the truth value given the quantifier and model description of sets A and B.

Van Bentham argues that the denotations of all first order quantifiers (defined in van Bentham, 1987) which satisfy the extensionality and conservativity constraints described above (and a third one, the quantity constraint[3]) can be computed using finite state automata with two internal states. These include the quantifiers *all, some, no,* and *not all.* For example, an automaton to evaluate the truth of *all of the squares are red,* reads in a string of symbols where '1' stands for those squares which are red, and '0' stands for those squares which are not red. The automaton accepts all occurrences of '1' and remains in a 'true' state, but on encountering a '0' it switches into a 'false' state which cannot be altered by further symbols. This captures the fact that only one non-red instance of a square is sufficient to falsify the sentence. Similarly, the automaton for *some* enters and remains in a 'true' state on encountering a '1' in its input.

Higher order quantifiers, such as most, *almost all, many, few* and *hardly any* cannot be computed by finite-state automata, but require the more complex push-down automata. As an example, consider van Bentham's description of *most* as a push-down automaton. As the input list of 1's and 0's is read, they

---

[3] The quantity constraint states that the truth of the quantified relationship only depends on the number of members of the model, noun set, predicate set and intersection between the noun set and predicate set, and not the identity of those members. So, given the statement that *some of squares are red,* then it does not matter which of the full set of squares are red, only that there is enough to determine that *some* is true.

are stored on the stack, which can only be read to a depth of one item. When a 1 is read off the input tape, it is compared with the top item on the stack. If this is a 0, then it is erased from the stack and the automaton reads the next input from the tape. Also, when the current input is 0 and the top item on the stack is 1, that top item is erased. Inputs are only added to the stack when they are the same as the current top stack item. The sentence is true if, at the end of processing, there are only 1's in the stack (this can be checked by reading and erasing them one at a time).

Van Bentham claims that the other higher-order quantifiers can be modelled by different permutations of push-down automata, but also argues that quantifiers like *many*, *few* and *almost all* are under-determined and must be given approximate values which retain the spirit of their meaning. He suggests, for example, that *many* can be interpreted as *at least one third,* and *almost all* is paraphrased as *at least two thirds.* However, despite the mathematical elegance of this approach, the inadequacy of such approximations should be obvious as an attempt to bypass the context-dependency of these quantifiers.

Psychological theories of processing complexity

While existing psychological theories provide a less detailed processing account than that suggested by either Barwise and Cooper or van Bentham, they do describe differences in verification time for sentences of positive and negative affect. Researchers have employed two general techniques to investigate the verification of negated or quantified sentences. Subjects are required to either evaluate the sentential description against an external model, generally a pictorial representation (Clark and Chase, 1972; Clark, 1976; Just and Carpenter, 1971; Carpenter and Just, 1975), or against commonly known semantic knowledge (eg. Greene, 1970; Fischler, Bloom, Childers, Roucos and Perry, 1983). Just (1974) reports that cross-modal verification and the comparison of sentences with semantic knowledge produce the same pattern of results, and argues that common processes are used in the two tasks (but see Tanenhaus, Caroll and Bever, 1976, for a criticism of the sentence-picture verification task).

Both Just and Carpenter, and Clark and his colleagues developed componential models of sentence and picture representation and comparison to explain results on the first of these tasks. Although there are some differences between the two accounts, there is general agreement on the core components. They argue for a four-stage process, where the sentence is first of all represented as a set of propositions, and then the picture is encoded as a similar set of propositions. Negatives, whether sentential negatives of the form *it isn't true that the dots are red* or predicate negatives like *it's true that the dots aren't red,* are treated as denials of a positive proposition that *the dots are red.* However, Just and Carpenter differ from Clark over the precise representational form for this, while Tanenhaus et al are critical of inconsistency in both theories. In general, the negated sentence is represented as an embedded affirmative proposition (what Clark and Clark, 1977, call the supposition) and an external negative, with the form NEG (DOTS, RED). Once the sentence and picture have been encoded as sets of propositions they are compared.

The comparison process starts from the assumption that the sentence is true, then attempts to match propositions from the sentence with others from the picture. Embedded propositions (or suppositions) are compared first. If these do not match, then the truth index of the sentence is changed from true to false. The embedding propositions are compared next and, again if these do not match the truth index is changed. So the index is changed if the sentence is prefixed with a negative marker, and the picture is affirmative. Each of these steps is assumed to add a constant increment to the verification latency, so a change in truth index following a comparison of both the embedded and embedding propositions will produce the longest verification latency. Also, because the representation of a negative takes an extra step - adding the embedding proposition, it is assumed to add more time to the verification latency.

Clark (1976) tested this theory by comparing negative and affirmative sentences. He recorded the time taken by subjects to verify the truth of sentences like those illustrated in (72) against a picture of a square, and reports response latencies with the order given in (72), where *true affirmative* sentences took the least time to verify, and *true negative* sentences took the most time. This pattern of results matches the model predictions. *True affirmatives* should take the least time because they require a simple unembedded propositional representation, and find a straightforward match with the picture representation. *False affirmatives* have an extra stage where where no match is established and the truth index is changed. Both of the negative stages imply more complex representational stages which add the embedding negative marker, and an additional comparison stage to evaluate that embedding proposition. The *true negative* has the longest verification latency because the truth index must be changed twice, once because the embedded proposition is false, and a second time because of the embedding proposition.

(72)	*True affirmative:*	The square is present
	False affirmative:	The circle is present
	False negative:	The square isn't present
	True negative:	The circle isn't present

Another study by Just and Carpenter (1971) established that the implicit negative components of quantifiers like *few* imply a similar increase in verification latency as explicit negative markers. They compared explicit negative quantification like (73) with comparable affirmatives like (74), implicit negative quantification (75) with quantified affirmatives (76), and what they called 'semantic negatives' (77) and the corresponding affirmatives (78).

(73) None of the dots are red.
(74) All of the dots are red

(75) Few of the dots are red.
(76) Many of the dots are red.

(77) A minority of the dots are red.
(78) A majority of the dots are red.

A longer processing latency for negatives was observed across all three categories. Explicit negatives were evaluated fastest, then the implicit negatives, with the 'semantic negatives' slowest. There was also evidence of the same pattern of results found by Clark, where true affirmatives were evaluated faster than false affirmatives, which in turn were faster than false negatives, and true negatives had the longest verification latency.

It is interesting to note that Just and Carpenter's so-called 'semantic negative' *a minority* shares many of the semantic properties of *only a few*. By my intuitions, it fails tests of upwards monotonicity, and is questionable on tests of downwards monotonicity. It is also unacceptable when tested for persistence, and borderline on tests of anti-persistence. Also, like *only a few* it is more acceptable with positive tags tests than those for negatives, however it licenses the use of negative polarity items. Compare (79) with (80).

(79) A minority of Chrysler dealers ever sell cars.
(80) A majority of Chrysler dealers ever sell cars.

In a subsequent study of eye fixations during a sentence-picture verification task, Carpenter and Just (1972) observed different strategies in the verification of quantifiers like *few*, and other quantifiers like *a minority of*. After reading a sentence like (75), subjects fixated on the set of non-red dots when comparing the sentence against a picture representation, but fixated on the set of red dots when verifying (77). Carpenter and Just interpret these results as evidence of the implicit negativity of *few*, but they also accord with the hypothesis-testing strategy earlier associated with the verification of monotone-decreasing quantified sentences. It was suggested that subjects attempt to disprove monotone-decreasing quantified sentences, but look for supporting evidence when verifying monotone-increasing ones.

More recently, Laport (1994) carried out a sentence-picture verification study on quantified sentences where she manipulated the monotonicity type. Subjects read a series of sentences relating geometric shapes to colours under monotone-increasing, monotone-decreasing and non-monotonic forms of quantification. Sentence (78) is an example of the monotone-increasing condition, while (79) is a monotone-decreasing example, and (80) describes non-monotonic quantification[4].

(78) All triangles are blue.
(79) No triangle is blue.
(80) Four triangles are blue.

Laport measured the reading time for these sentences, and the time for their verification against a picture of coloured geometric shapes. She observed increased reading time for both monotone-decreasing and non-monotonic quantification relative to the monotone-increasing condition. She also observed a substantial increase in verification time following monotone-decreasing quantification compared to monotone-increasing quantification. However, the difference was less marked for monotone-increasing compared to non-monotonic quantification, although the error rate was four times as great following non-monotonic quantification. In general the results confirm Barwise and Cooper's prediction, but without finding any added complexity for the processing of non-monotonic quantifiers beyond that of monotone-decreasing ones. However, the error rates do suggest greater difficulty in evaluating non-monotonic quantified sentences, which might explain the surprising low verification time. It may be that subjects adopted a different strategy in response to the added difficulty.

---

[4] The sentences were presented in Dutch. The translation is Laport's own.

It is worth noting the similarity between these theories and the 'Derivational theory of complexity' (c.f. Miller and Isard, 1963), which argued that the transformational complexity of a sentence is directly related to processing complexity. More specifically, the theory claimed that the more transformations needed to map the surface syntactic structure of a sentence onto its underlying structure (where both of these are defined in Chomsky, 1957), then the greater the difficulty in processing this sentence. Both Clark, and Just and Carpenter, suppose that the processing complexity of a negative sentence is directly related to the number of steps needed However, the 'Derivational theory of complexity' was largely abandoned because it was not clear that the evidence of increased processing times for particular sentences was actually due to transformational complexity, and at the same time there was counter-evidence that transformational complexity did influence processing (c.f. Fodor, Bever and Garrett, 1974).

The first of these two criticisms can also be levelled at theories of linguistic processing based on the sentence-picture verification methodologies. In particular, it is not clear if the latency differences reported by both Clark, and Just and Carpenter, are due to processing differences during sentence comprehension, or when comparing the sentence against the picture. This is a major criticism, which implies that this is not an appropriate methodology for investigating the early stages of sentence comprehension.

Conclusion

This chapter has reviewed some model-theoretic and psychological approaches to natural language quantification. It provided an overview of model-theoretic semantics, and argued for a correspondence between this and psychological approaches to semantics. More specifically, it was argued that model-theoretic semantics provides an account of linguistic competence, while psychological approaches must specify linguistic performance.

Given this parallel between model-theoretic and psychological semantics, the chapter went on to consider model-theoretic accounts of natural language quantification, particularly those accounts proposed by Montague (1974), and researchers influenced by generalised quantifier theory (Barwise and Cooper, 1981; Keenan and Stavi, 1986; Westerståhl, 1989). This concentrated on attempts to categorise quantifiers in terms of semantic properties. It was observed that failure to observe the semantic constraints of conservativity and extensionality implied that the interpretation of a quantifier is context-dependent. Other semantic properties, i.e. monotonicity and persistence, describe patterns of directional inference associated with quantifiers. Those quantifiers which are monotone-increasing according to linguistic tests, are also negative. That is, they are either explicitly marked as negative (e.g. *not many*), or are implicitly negative (e.g. *few*).

The chapter then described three accounts of the processing of natural language quantifiers. The first of these accounts (Barwise and Cooper, 1981) claimed that the semantic properties of monotonicity and persistence have consequences for the processing of quantifiers. They argue that monotone-decreasing quantifiers like *few* and *not many* have a greater processing complexity than monotone-increasing ones like *a few* and *many*, while non-monotonic quantifiers like *only a few* are more complex yet. Moreover, quantifiers of the same monotone processing complexity can be separated in terms of the property of persistence. Anti-persistent quantifiers are more complex than persistent quantifiers which have the same monotone complexity. It was shown that although the account of processing complexity is not entirely satisfactory, it does produce evidence that monotone-increasing and -decreasing quantified sentences are processed in different ways. The second account (van Bentham, 1986) proposed that the processing of quantifiers could be described for in terms of the complexity of the semantic automaton required to evaluate the truth of a quantified sentence. The third account (Just and Carpenter, 1971; Clark and Chase, 1972) was a psychological model of decision latencies to evaluate the truth of quantified sentences relative to pictures. While the theories and methodology are flawed, there was some evidence that negative, and therefore monotone-decreasing, quantifiers are more complex than positive (monotone-decreasing) ones. This was further evidence that there are processing consequences of monotonicity.

The next chapter argues that the monotonicity of quantifiers correlates a psychological property of focus that is described by Moxey and Sanford. It will be shown that monotone-increasing and -decreasing quantifiers exhibit different patterns of focus, and that this has consequences for the interpretation of subsequent pronominal reference.

Chapter two:

The interpretation of quantifiers

Introduction

The previous chapter described some formal approaches to semantics, particularly the semantics of quantification, and suggested that there may be a parallel between the formal classification of quantifiers on the basis of either negativity or monotone inference patterns, and the psychological account of quantifiers proposed by Moxey and Sanford (1987; 1993a). This chapter will detail Moxey and Sanford's theory and experimental findings, alongside some related research on the interpretation of quantifiers and quantificational adverbs.

Moxey and Sanford argue that there are two classes of quantifier which are used to divide noun-phrases (NPs) into **refset** and **compset** partitions, that these quantifiers exhibit differential patterns of focus, and that this is indexed by the ease of pronominal reference to each of the NP partitions. Those quantifiers like *a few* and *many* which appear to be positive and monotone-increasing according to formal tests, focus on and enable reference to the refset partition, the NP subset of which the VP predicate is true. This means that plural pronominal reference to a sentence like (1) should be interpreted as reference to those MPs who attended the meeting.

(1) A few of the MPs went to the meeting.

In contrast, those quantifiers which appear to be negative and monotone-decreasing, such as *few* or *not many,* also focus processing attention onto the compset partition, or the NP subset of which the VP predicate is false, and permit reference to either partition. For example, plural pronominal reference to a sentence like (2) can refer to either those MPs who were present, or those who were absent from the meeting.

(2) Few of the children went to the park.

It seems then that the two classes of quantifier are used to control **discourse focus** in different ways. Monotone-increasing ones place the refset in discourse focus, but monotone-decreasing ones are used to place both subsets in discourse focus.

Discourse focus and quantification

Before presenting the evidence that quantifiers are used to differentially focus on and enable reference to subsets of the quantified NP, I will clarify what is meant by focus, and one consequence of this account of quantification.

The notion of discourse focus has become common in theories of discourse processing and representation (cf. Garnham et al, 1989; Greene, McKoon and Ratcliff, 1992; Garrod, Freudenthal and Boyle, 1994), but is most clearly formulated in Sanford and Garrod's (1981; Garrod and Sanford, 1982, 1991) scenario theory of text comprehension. They describe a two component mental representation of text. The main part, called **explicit focus,** is a representation of the objects and relations described by the text. A second partition (called **implicit focus**) maps this explicit focus representation onto an underlying description of the situation (called a **scenario**) and the appropriate discourse roles of entities in explicit focus. This allows situation-specific knowledge to flesh out the details of the impoverished text input.

Following Chafe (1972, 1976), the objects and relationships in explicit focus are focal (or **foregrounded,** to use Chafe's terminology), and only these elements are accessible to pronominal reference. For example, Garrod and Sanford (1982) used discourse fragments like (3) and (4) to demonstrate the incapability of pronominal reference to objects in implicit focus. They argue that although a car forms part of the scenario description of both sentences, as demonstrated by successful definite noun-phrase reference in (4), pronominal reference fails in (3) because the car is an implicit entity which has been cued by the verb *to drive,* and resides in implicit focus.

(3) Keith was driving to London.
It had recently been overhauled.

(4) Keith was driving to London.
The car had recently been overhauled.

The account presupposes that there are just two discrete levels of focus, while Chafe also resists the suggestion that foregrounding might be a continuous phenomenon where potential antecedents exhibit degrees of accessibility (see discussion section in Chafe, 1972).

The claim that monotone-decreasing quantification focuses on and enables reference to the compset is problematic to Sanford and Garrod's theory, because it violates the requirement that only those entities which are described in the text can be represented in explicit focus. By its very definition, the compset is not an explicitly described entity, but the complement set of that described by the quantified sentence. Yet it is accessible to pronominal reference, which implies that it must reside in explicit focus. The solution to this problem (Garrod and Sanford, 1991) is to allow explicit focus to include those entities which are introduced by semantic operators like monotone-decreasing quantifiers, and maintain implicit focus as a representation of the relevant situation-specific knowledge which has been retrieved from long-term memory.

This solution also means that focus cannot be discrete, since I have claimed that both the refset and compset are plausible antecedents of pronominal reference following monotone-decreasing quantification, although the compset is the preferred antecedent. There must instead be at least two levels to the explicit focus partition, one which holds the more focal compset, and another for the refset. This can be generalised to claim that entities in explicit focus (and presumably also those in implicit focus) exhibit degrees of accessibility. This might be represented as an ordered list using a symbolic processing paradigm, or as a range of activation values in a connectionist framework. The latter account supposes that these are represented in parallel, but that the compset is generally more accessible than the refset. However, only the refset may be represented and open to reference following monotone-increasing quantification.

Focus can also be thought of as a form of perspective control where the currently focal element takes charge of the narrative direction. Other instances of perspective controllers are contrasting bipolar descriptions, and perhaps also the scalar expressions investigated by Fodor (1975), Ladusaw (1979), and Fauconnier (1978). One pair of contrasting bi-polar descriptions are the quantificational adjectives *half-full* and *half-empty*. If a river at 50% capacity is described as *half-full* or *half-empty* then, although the two descriptions are semantically equivalent, they hold different implications. Describing the river as *half-full* directs attention towards the amount of water which is present, but to describe it as *half-empty* focuses on the missing water[5]. This may also constrain the type of inferences which are made.

Experimental evidence of quantifier focus

Moxey and Sanford (1987) used an off-line sentence-continuation task to investigate their claim that quantifiers focus on and enable reference to different NP subsets, and to show that there is a systematic variance in the content of inferences. Subjects were asked to provide a one or two sentence continuations to discourse fragments like (5), where the quantifier was one of *a few, few, many, not many* or *only a few*. It was predicted that the plural pronoun *they* would be interpreted as referring to the focal NP subset, and that it would be possible to determine which subset was referred to from the content of the continuations. Figure 1 summarises for each of the quantifiers the percentage of continuations which described the compset.

(5) QUANTIFIER of the MPs attended the meeting. They . . .

QUANTIFIERS	REFSET CONTINUATIONS (%)	COMPSET CONTINUATIONS (%)	OTHER CONTINUATIONS (%)
A few	95	0	5
Only a few	95	5	0
Few	35	62.5	2.5

Table 1: Percentage frequency of refset, compset and other intended referents of plural pronoun continuations in different quantificational conditions (from Moxey and Sanford, 1987).

Moxey and Sanford found that the plural pronoun was almost always interpreted as reference to the refset partition of the quantified NP following quantification by *a few,* and interpreted this as evidence that the quantifier focuses processing attention onto this set. Similarly, the continuations almost always referred to the refset following quantification by *only a few* [6], with a small incidence of compset reference. Again this was interpreted as evidence for refset focus. However, there was a more diffuse pattern of results following quantification by *few,* which suggested that the pronoun could be interpreted as reference to either the refset or compset partition of the quantified NP, although the results also suggested that there was a preference for compset

---

[5] This has an interesting folk psychology interpretation as an indicator of character type, and features in a Gary Larsen cartoon. An optimist is believed to see things from a positive perspective, so the glass is always *half-full,* while the pessimist takes a negative view and the glass is disappointingly *half-empty.*

[6] A more detailed account of the experimental results for *only a few* is presented in Chapter 8.

reference. Moxey and Sanford concluded that quantification by *few* is used to place the compset in focus, but that the refset remains a plausible referent.

A second study directly compared the frequency of continuations which focused on the compset following quantification by either *few* (61%) or *not many* (79%). Moxey and Sanford concluded that quantification by *not many* also results in compset focus, but more reliably than *few*, perhaps because of its explicit negativity.

Moxey and Sanford also found differences in the content of subject's continuations which correlated with the focusing properties of the quantifiers. They found that continuations to sentences quantified by *a few* or *only a few* generally extended the narrative from the perspective of the refset (eg 6) - Moxey and Sanford refer to these as "what happens next" continuations - and to a lesser extent gave reasons why the refset is true of the predicate (eg 7).

(6) Many of the MPs attended the meeting. They discussed NHS funding levels.
(7) A few of the MPs attended the meeting. They were obliged to go by the party whips.

However, continuations to sentences quantified by *few* or *not many* tended to explain why the compset is not true of the predicate (eg 8).

(8) Few of the MPs went to the meeting. They thought that it would be boring.

The data supports the claims that there are two classes of quantifier which differentially focus on and enable reference to the refset and compset subsets of the quantified NP. Moreover, the results suggest that only the refset may be available as an antecedent following quantification by *a few*, but that both the refset and compset are possible antecedents following quantification by *few* and *not many*. Finally, a more detailed analysis of the sentence continuations suggests that the quantifiers are used to direct inferential processes. This results in simple narrative continuations following quantification by *a few*, but continuations which attempt to explain the state of events following quantification by *few* or *not many*.

Experimental evidence of context-dependent quantifier function

Given the experimental evidence that quantifiers *a few*, *few*, and *not many* are used to direct discourse focus, it remains to be seen whether they also exhibit the context-dependency claimed for them in Chapter one. We saw there that the non-extensionality and possible non-conservativity of these quantifiers implied that the truth of a sentence quantified by them depends on properties of the model other than those explicitly described in that sentence. For instance, researchers agree (Keenan and Stavi, 1986; Westerståhl, 1989) that the truth of a sentence quantified by either *few* or *many* requires a comparison of the actual sentence extension against the size of an expected set extension, although Keenan and Stavi argue that the expected set may be unknowable.

However, there are no empirical studies of these semantic observations, and experiment research has instead concentrated on the context-dependent mapping of vague expressions (generally frequency adverbs) to numbers (cf. Pepper and Prytulak 1974, Newstead and Collis, 1987). Newstead and Collis, for example, asked subjects to translate the frequency of events described by frequency adverbs like *never*, *rarely*, *often* or *always*. Such studies generally agree that there are contextual constraints on the mapping to numerical values for expressions which denote large amounts or frequencies, but no effect for expressions which denote a small amount or frequency. For instance, Newstead and Collis demonstrated that the higher the base-rate expectation, then the higher the value attributed to high-frequency denoting terms like *often* and *always*, while the numerical value for low-frequency denoting expressions like *never* and *rarely* remained stable under a contextual manipulation.

Other studies by Hörmann (1983) have investigated the mapping of quantifiers onto numbers. She asked subjects to assign numerical values to a number of quantified noun-phrases across a range of contexts, and identified several important variables. For instance, she found that certain types of objects, and those which were small, were rated as more numerous than others. One example of this was the finding that subjects produced larger numerical values to describe *many crumbs* than to describe *many mountains.* The spatial location of the objects also proved to be important, and subjects rated *a few* people standing in front of a building as more numerous than *a few* people in front of a hut. Hörmann interpreted this to mean that the size of the building affected the numerical interpretation of the quantifier. However, Newstead (1988) suggests that it might instead reflect the expected frequencies of occurrence, where subjects try to imagine how many is likely for the described situation.

To test this, Moxey, Tuffield and Temple (reported in Moxey and Sanford, 1993a) carried out a similar experiment where subjects were asked to assign numerical values to noun-phrases quantified by *a few* and *many* in sentences like (9) and (10). They established that it was not just the size of the building which determined the numerical mapping for *many*, but the expected frequency of people outside the building, which supports Newstead (1988). Because a chip shop would be busy with customers, a large number of people might be expected outside it, but people are not expected to crowd around a fire station to the same extent. As with previous studies no effect was observed for the small amount denoting quantifier *a few*.

(9) QUANTIFIER people standing in front of the fish and chip shop.
(10) QUANTIFIER people standing in front of the fire station.

A further experiment (Moxey and Sanford, 1993b) used a fully independent design to investigate the contextual constraints on numerical mapping for the quantifiers *few, very few, only a few, quite a few, not many, many, very many, quite a lot* and *a lot*; and again found that numerical values assigned to the high-amount denoting quantifiers varied with context, but no effects were observed for the other low-amount denoting quantifiers. Moxey and Sanford also observed a poor differentiation between the values assigned to quantifiers like *few* and those like *very few* which include intensifiers. However, there are strong focusing differences between these same quantifiers (Moxey and Sanford, 1987) which suggests that the intensifiers might have a discourse function rather than an influence on expression to number mapping.

The lack of a context-dependent effect for the low-amount denoting quantifiers is more troublesome, but may be due to floor effects. For example, if a contextual manipulation doubles the reported values across all of the quantifiers, then the difference for old and new values will not be so marked for a small quantifier like *a few* compared to a large quantifier like *many*. It might be the difference between 5% and 10% compared to a difference between 40% and 80%. Some support for this explanation is given by a study conducted by Moxey, Sanford and Grant (reported in Moxey and Sanford, 1993a), which did find a small but reliable context effect for small quantifiers when a broader range of contexts was used.

All of these experiments describe the mapping of vague expressions onto numerical values, yet Moxey and Sanford's sentence-continuation studies demonstrate that quantifiers have a discourse function where this mapping is a secondary process, if one that occurs at all. For present purposes, it may be more pertinent to look for context-dependency in the discourse function of quantifiers. For example, we noted earlier that compset focus is evident in both the type of reference and content of continuations produced by subjects. Continuations following quantification by *few* and *not many* tend to explain the smallness of the refset. Clearly, such inferences depend on knowledge about the described situation, and it is possible that compset focus is mediated by the availability of inferences in a given context, and explanations may be unwarranted under some conditions.

Moxey and Sanford (1987) manipulated the VP predicate of a sentence quantified by *few* to test this hypothesis. They asked subjects to provide continuations to the discourse fragments in (11) and (12). Fragment (11) matches the expectation that children are fond of Santa Claus, and only a very small refset would normally be expected. It seems likely that there will be few, if any, inferences which explain the small refset for this situation. Fragment (12) describes the contrasting situation where there is a small refset who do like Santa Claus, which violates expectations and will license a number of explanatory inferences.

(11) Few of the children hated Santa Claus. They . . .
(12) Few of the children liked Santa Claus. They . . .

As predicted there were more continuations which explained the small refset in the expectation-violating condition (80% compared to 30%), but there were no reliable differences in the tendency towards compset reference. Seventy-five percent of continuations described the compset in the expectation-violating condition, and 90% did in the expectation-matching condition. This invites the conclusion that compset focus is primarily under semantic control, and pragmatic factors only influence the content of continuations. However, a subsequent experiment (Moxey and Sanford, 1993a) demonstrates that the availability of knowledge-based explanatory inferences determines the likelihood of compset focus following quantification by *only a few.* One of the experimental materials (14) violated expectations, but there were no strong expectations associated with the other situation (13). Moxey and Sanford found that while two thirds of continuations for (14) referred to the compset, there were hardly any for (13).

(13) Only a few of the MPs were at the meeting. They . . .
(14) Only a few of the children ate their ice-cream. They . . .

This suggest that both semantics and pragmatics can contribute to the likelihood of compset focus, but that pragmatics will only predominate if the semantic constraints are weak. Quantification by *few* is a strong semantic cue to compset focus, so a manipulation of the context has a negligible effect. However, the non-monotonic *only a few* exerts much weaker semantic constraints, and focus is more strongly influenced by contextual manipulations.

Other evidence of perspective control by function words

I have suggested that other syntactic classes might also exhibit focus, and there is some evidence for this. Moxey, Sanford and Barton (1991), for instance, describe how the quantificational adverbs *seldom, rarely, occasionally* and *only occasionally* appear to exhibit focus preferences. Given a sentence like (14), these adverbs divide the set of occasions when Mary eats into a refset of those when she eats in a restaurant, and a compset of when she eats elsewhere[7]. The adverbs *seldom* and *rarely* appear to focus on the compset partition and license a continuation sentence which explains why Mary does not eat in restaurants (15), while a sentence which refers to the refset (16) appears anomalous. In contrast, quantification by *occasionally* focuses on the refset and favours continuation by (16). Finally, *only occasionally* does not exhibit any strong focus preferences, and both continuation sentences appear equally acceptable.

(14) Mary ADVERB eats in restaurants.
(15) She enjoys eating at home.
(16) She likes to be waited on.

Moxey et al demonstrated these preferences experimentally. They asked subjects to complete the text fragments in (17) and (18) for each of the four adverbs, and classified the completed continuations by their content. They found that *occasionally* predominantly resulted in continuations which gave reasons for going to the cinema, while *seldom* and, to a lesser extent, *rarely* produced reasons for not going. Both sets of effects were made stronger when the connective *because* was used to link the quantified and anaphoric sentences (17). *Only occasionally* showed no strong preferences, although conjunction by *because* enhanced the frequency of reasons for not going to the cinema.

(17) John ADVERB goes to the cinema. He . . .

[7] It should be noted that there is a scope ambiguity associated with the quantificational adverbs, and they could have a restricted scope where they focus on the frequency with which Mary eats as opposed to some other act that might be performed in a restaurant. For instance Mary might be a waitress who often works, but rarely eats in restaurants.

(18) John ADVERB goes to the cinema because he . . .

Moxey and Sanford interpret the results for *rarely* and *seldom* as evidence that they are adverbial correlates of *few* and *very few* respectively; while the function of *occasionally* appears to parallel that of *a few*, and *only occasionally* behaves similarly to *only a few*.

Moxey and Sanford are not the only researchers to observe perspective control by function words. Anscombre and Ducrot (1986) present a linguistic account of the opposed function of the French mass quantifiers *un peu* and *peu*, which are equivalent to the English quantifiers *a little* and *little*. They argue that the quantifiers are used to 'bring out different facts' about utterances in which they appear. For example, (20) carries the implication that the quantity of work is small, while (19) simply asserts that some work has been done. Some of Anscombre and Ducrot's theoretical claims have been empirically tested. Champaud and Bassano (1987), for example, report a developmental study which investigates the contrasting perspectival function of adverbs like *seulement* (*only*) and *bien* (*really*, or *at least*).

(19) Peter has done a little work.
(20) Peter has done little work.
(adapted from Anscombre and Ducrot, 1986)

Another experimental study by Jarvella and Lundquist (1994) investigated the effect of Danish and Finnish scalar expressions on the processing of co-reference within a discourse. They observed that the ascending scalarity of *almost*, and the descending scalarity of *only* interact with pragmatic knowledge about the described situation. For instance, they found that these quantifiers interact with the knowledge that 'the more votes someone gets, the more likely he is to win' in sentences (21) to (24). The quantifier *almost* is used to mark the fact that a large number of votes were gained relative to those needed to win the contest. This licenses the inference that John Smith is likely to win, and enables a co-referential interpretation in (21) but not (22). In contrast, the quantifier only is used to mark the fact that a small number of votes were gained relative to those needed to win, which enables a co-referential interpretation in (24) but not (23).

(21) John Smith got almost 500 votes in the first round. He is likely to win the election.

(22)?? John Smith got almost 500 votes in the first round. He is likely to lose the election.

(23)?? John Smith got only 500 votes in the first round. He is likely to win the election.

(24) John Smith got only 500 votes in the first round. He is likely to lose the election.

Jarvella and Lundquist tested these claims in several experimental studies where subjects read experimental passages like (25), which described the performance of two named characters in a competitive scenario, and consisted of a context-setting section, followed by two sentences. The first of these sentences contained a statement about the performance of one of the characters (who was described by name) and included either an ascending or descending scalar expression. These were the quantifiers *almost all* or *only a few* respectively. The second sentence began with a noun-phrase (*the tall blond Århus-girl* and *the 20-year-old office girl*) which referred indeterminately to one of the passage characters, and indicated that this person was currently either winning or losing the competition (*behind / ahead*).

(25) The Beauty Competition

The year's big beauty competition took place in the Circus Theatre in Copenhagen. Eighteen models participated, chosen from various provincial cities. As journalists and other media people had suspected. Karina Madsen and Louise Fernholm stood out as the big favourites.

s1 In the round with bathing suits, Karina Madsen received high marks from *only a few / almost all* the judges.

s2 The tall blond Århus-girl was clearly *behind / ahead.*

Jarvella and Lundquist predicted that the anaphoric NP would only be interpreted as co-referential with the previously named character when the scalar expressions of the quantified and anaphoric sentence ran in the same direction. This meant that the *tall blond Århus-girl* would only be interpreted as reference to *Karina Madsen* when she received high marks from *almost all* of the judges, and the anaphoric sentence described her as being *ahead.* Similarly, co-reference would be possible when *only a few* and *behind* were used together.

This prediction was supported in an off-line study where subjects indicated which character was in a winning and losing position. There was 95% agreement that the immediately preceding named character was the appropriate antecedent of the anaphoric NP when the scalar expressions were in the same direction, and 80% agreement that the other named character was the antecedent when the scales were in opposite directions. These preferences were also evident in a self-paced reading experiment where YES / NO responses to questions of the form "Did N win?" or "Did N lose?" achieved more than 80% responses in the predicted directions. Jarvella and Lundquist also found that there was an increased reading time for sentences containing downwards scalar expressions, which suggests that these produced greater processing difficulty than upwards scalar expressions.

It seems then that there is considerable evidence of perspective control by function words, and several syntactic classes readily decompose into subsets of operators which exhibit contrasting patterns of focus. It is my contention that these function words are used to control the structure of mental representations and direct inference processing during comprehension. This requires that function words like quantifiers and quantificational adverbs directly correspond to mental procedures, which are triggered on reading the function word. However, this is a contentious claim because none of the studies described here are tests of perspective control during reading. The experimental chapters test a more specific version of this claim, that the success of anaphoric reference to a quantified NP is contingent on the focus patterns of the quantifier. But before presenting the results of these studies, I want to sketch some representational requirements, and in the next chapter I will consider the likely processing implications of this account.

Quantifiers and Mental Models

My starting point is Johnson-Laird's theory of Mental Models (Johnson-Laird, 1983; Johnson-Laird and Byrne, 1991). This was introduced in the previous chapter as an account of mental representations which was in the spirit of model-theoretic semantics, but with an awareness of the performance restrictions and constructive processes required by a psychological implementation. It is intended to subsume discourse representation within a general account of reasoning. However, quantifiers are only considered for their contribution to reasoning, particularly syllogistic reasoning (eg. Steedman and Johnson-Laird, 1978; Johnson-Laird, 1983; Johnson-Laird, Byrne and Tabossi, 1989; Johnson-Laird and Byrne, 1991), and are absent from the account of discourse comprehension. This is unfortunate because there are good reasons to suppose that the inferences required for complex tasks like syllogistic reasoning are quite different from those required for comprehension (cf. Fodor, 1983; Graesser, Singer and Trabasso, 1994, Kintsch, 1994), and it is at least conceivable that the representations will also differ. This distinction is central to the debate about automaticity and the time course of inferences reviewed in the following chapters.

The Mental Models approach to quantifiers has other failings too. Johnson-Laird does not venture far beyond traditional logical quantifiers like *all, some* and *none* (Bach, 1993). He does includes *most, more than half* (Johnson-Laird, 1983, p137-141) and *only* (Johnson-Laird and Byrne, 1991), acknowledges J. D. Fodor's similarly model-based approach to the quantifiers *each* and *every* (Fodor, 1982)[8], and assumes that the theory will also apply to *many, several* and *a few* (ibid., p137), but without specifying how. Neither does he include any of the semantic properties reported by generalised quantifier theory, or include focus within his models. Other mental models theorists have added focus and offer a broader and more psycholinguistic account of mental models (eg Garnham et al, 1989), but do not discuss quantification. Despite these weakness, Mental Models make some assumptions about the representation of noun-phrases and function of quantifiers which can kick-start my own account.

According to Johnson-Laird, a Mental Model is an internalised tableau representation which is constructed from the semantic content of language and general knowledge about the described situation. Plural nouns are represented in these models by an arbitrary number of tokens, except in cases involving numerical quantifiers and other forms of number restriction (eg. *four dogs, both donkeys*), which stand for instances of the noun type; and quantified sentences are represented by the necessary and possible subsets of the quantified noun relative to its predicate.

For example, a sentence of the form *All X are Y* requires that the X terms are exhaustively represented in a one-to-one relation with Y. The brackets around the X terms in Figure 1 (adapted from Johnson-Laird and Byrne, 1991) assert that the set of X's has been fully represented, and if the model was fully 'fleshed-out' to include all instances of X, then these would necessarily stand in relation to a Y term. In the original theory, other Y terms are included

[8] Fodor (1982) is detailed in Chapter 9. She distinguished between *all, each* and *every* using the distributive property discussed there.

separately from the X terms to allow for the possibility of other instances of Y which are not X.

(X)	Y
(X)	Y
(X)	Y
(X)	Y

Figure 1: Mental models representation of 'All X are Y', from Johnson-Laird and Byrne, 1991.

In the same way, *some X are Y* is represented by an arbitrary number of X that stand in relation to Y, and the possibility that there are other instances of both X and Y which do not relate to each other. *No X are Y* is described by the necessary condition that X and Y are exhaustively represented, and no X and Y stand in relation to each other. Finally, *some X are not Y* is represented by an arbitrary set of X standing in relation to an exhaustively represented set of Y's, with the addendum that the set of 'not X' is exhaustively represented as 'not Y'.

Johnson-Laird argues that mental models also cope with quantifiers denoting relative set sizes, such as *most, many several* and *a few*; and demonstrate inferences which are dependent on these relative sizes and an interaction with general knowledge. He illustrates this with the following examples (26 and 27).

(26) All fascists are authoritarian.
Most authoritarians are dogmatic.

(27) All archbishops are Tories.
Most Tories are middle-class.

The two premises in (26) invite the conclusion that *Most fascists are dogmatic,* but the same form of conclusion, that *Most archbishops are middle-class,* is blocked in (27) by the knowledge that the set of archbishops is very small compared to the set of middle-class people.

While there is some question about whether individuals represent the necessary and possible conditions of quantifiers in the way Johnson-Laird claims (eg. Newstead, 1989), and the use of arbitrary numbers of tokens means that it is difficult to separate the different representations for quantifiers like a *few, some, several* from each other and from numerical quantifiers, a more basic problem is that this form of representation is only suitable for some types of plural nouns and quantifiers. Others, like mass nouns (cf. McCawley, 1981) or group nouns (Barker, 1994), do not readily decompose into member elements in the manner assumed by mental models. For example, it is not possible to decompose *water, furniture* or *prose* into their component elements without losing the defining properties of the plural NP. This mean that the elements of water are not themselves described as water; and while water is wet, it is not so clear that a molecule of water is wet.

There are other differences too. For instance, adjectives which can apply to a mass noun are often not applicable to the individual elements. So you can have *a cup of dirty water*, but not *a molecule of dirty water*. Similarly, although the members of a count noun can be enumerated using the indefinite article or numerical quantifiers (eg. *three students, a house, more than four games*), and count nouns accept quantification by *a few, several, many* and *each*, this is not possible with mass nouns. In fact, they reveal their mass representation through the acceptability of reference by the neutral pronoun *it*[9] (see 28 and 29); and can be quantified by the likes of *a little* and *much* which appear anomalous when applied to count nouns.

(28) The furniture was broken. It was thrown out with the refuse.
(29) Some of the prose was obscene. It was removed by the editor.

Barker argues that group nouns like *committee* or *army* must be represented in manner such that the plural noun holds properties and licenses inferences independently of its underlying membership. What this means can be demonstrated using the group noun *committee* if we imagine a situation where the members of a committee are specified in the discourse - let's say that Tom, Bill and Harry are the only members of this committee. Barker argues that (30), (31) and (32) could all be statements about this situation.

(30) Tom, Bill and Harry discussed the best way to deal with joy-riders.
(31) The men discussed the best way to deal with joy-riders.
(32) The committee discussed the best way to deal with joy-riders.

However, while these appear to describe the same situation, that is, they share the same reference, they are not semantic equivalents. The truth of (32) entails

---

[9] However some forms of quantification are referred to using a plural pronoun, but this seems to have more to do with the form of quantification than the mass noun.

that (31) and (32) are also true, but neither the truth of (30) nor (31) entail the truth of (32). It could be that the three members of the committee are having a discussion outside of the situation where they constitute that committee.

Given that neither mass or group nouns cannot receive an individuated representation, I want to suggest that quantified count nouns do not receive it either. In fact, it is the form of quantification itself which is used to mark the granularity of the representation. This assumes that plural NPs have an unindividuated representation as a default, meaning that the members of the NP are not represented by individual tokens, and that some forms of quantification, say by *a few* or *few*, are used to produce a more fine-grained representation. As already described, these divide the NP into refset and compset partitions. The subsets will also have an unindividuated representation.

Other quantifiers with distributive properties (cf. Fodor, 1982), such as *each*, may be used to produce an even more fine-grained representation of the quantified noun. This may be fully individuated in the case of *each*, which would be consonant with Vendler's (1967) characterisation of it as a procedure which considers the member of set one by one. It is also consonant with the observation that distributive quantifiers like *each* and *several* only operate on count nouns (McCawley, 1981), since mass nouns cannot be individuated.

This approach can fit within the Sanford and Garrod's (1981; Garrod and Sanford, 1991) theory of discourse representation. It requires that plural NPs are represented as a unindividuated entity in Explicit focus, and map onto a single discourse role in Implicit focus. This may be marked as a complex representation in order to allow it to accept plural pronoun reference. Quantification by *a few* or *few* will divide the NP into refset and compset partitions, which are mapped into separate discourse roles, and licences pronominal references to one or other of the subsets, depending on the form of quantification[10]. Finally, quantification by a distributive quantifier, say *each*, will produce an even more individuated representation where a number of NP set members are mapped into separate roles in Implicit focus. This licenses singular pronoun reference to the individual members (33), and may be used to mark distinctions between the members of the NP set. For instance, my interpretation of (33) is one where the members of a group of MPs attend

---

[10] It is also possible that monotone-increasing quantification, say by *a few*, will only result in a representation of the refset.

separate meetings. More will be said in Chapter 9 about the interpretation of this type of sentence.

(33) Each of the MPs attended a meeting, and she took careful notes.

Since plural reference to the superset is always an option following both types of quantification, it seems likely that the representations are hierarchically organised, or embedded within each other. This means that representations of a quantified sentence are simply more fine-grained versions of the representation of an unquantified sentence.

Conclusion

This chapter has described some evidence from off-line production experiments that those quantifiers that monotone-increasing and -decreasing quantifiers exhibit different patterns of focus. The data can be interpreted as evidence that monotone-increasing quantifiers like *a few* and *many* are used to focus processing attention onto the refset of a quantified sentence, i.e. the set for which the sentence is true. Plural pronoun reference to this sentence is almost always interpreted as reference to the refset. In contrast, it appears that monotone-decreasing quantifiers like *few* and *not many* are used to place the compset in focus, where the compset is one of those sets of which the sentence is false. The experimental data can be interpreted as evidence that although both the refset and compset are plausible referents of plural pronoun reference, the compset is a preferred referents.

The chapter reviewed some other experimental evidence that the interpretation of non-logical quantifiers, like *a few*, *few*, *many* and *not many*, is context-dependent. It also reviewed some further experimental work by Jarvella and Lundquist (1994) which argues that positive and negative quantifiers are used to direct processing attention. Finally, the chapter considered the mental representation of quantified information, and argued that quantification is used to control the granularity of mental representation of sentence meaning.

However, the main aim of the chapter was to provide experimental evidence for the claim that quantifiers can be differentiated in terms of focus. Yet the reported experiments used off-line production tasks in which subjects provided completions to sentence which referred to a quantified sentence. Because they are production tasks, and because they are not time-limited, they may not describe processing during comprehension. For instance, it is possible that the observed effects only occur when subjects have time to deliberate on the appropriate interpretation of the pronoun. For this reason, the following chapter describes an account of the processing of reference to a quantified sentence during comprehension, and reviews previous research on the processing of reference and inference during comprehension.

Chapter three:

Processing anaphora and quantification during reading

Introduction

The previous chapter reported some sentence-continuation data in support of the claim that quantifiers are used to focus on and enable reference to particular NP subsets, and made some further claims about how these quantified NPs might be mentally represented. These claims will be transformed into some more specific processing predictions. In order to do so, I will isolate a set of likely component processes, then assess the compatibility of these with current models of text comprehension and inference processing.

The component processes of anaphoric reference

The three predicted components of pronominal reference to quantified noun-phrases are illustrated in Figure 1. First of all, the quantifier divides the noun set into refset and compset partitions, of which the sentence predicate is respectively true or false. However, it remains open as to whether this happens during the processing of the quantified sentence, or in response to the anaphoric reference. While this question is not addressed by the current studies, a proposed experimental manipulation is described in Chapter 6. Yet, whatever the locus of quantifier function might be, the other processes cannot begin before encountering the anaphor. These are both stages of anaphoric reference, where the anaphor is initially attached to the most focal antecedent, then this attachment is evaluated, and re-assigned if it proves to be anomalous.

The first stage of anaphoric reference, when the anaphor is initially attached to the most focal element of Explicit focus, is a heuristic which usually avoids the greater processing load involved in directly comparing the anaphor against all possible antecedents. It is based on the assumption that those entities which are in focus are also central to the current discourse, and the most likely recipients of pronominal reference. Other forms of reference, such as definite NPs, can be used to access less focal antecedents (eg Sanford and Garrod, 1981; Garrod and Sanford, 1991).

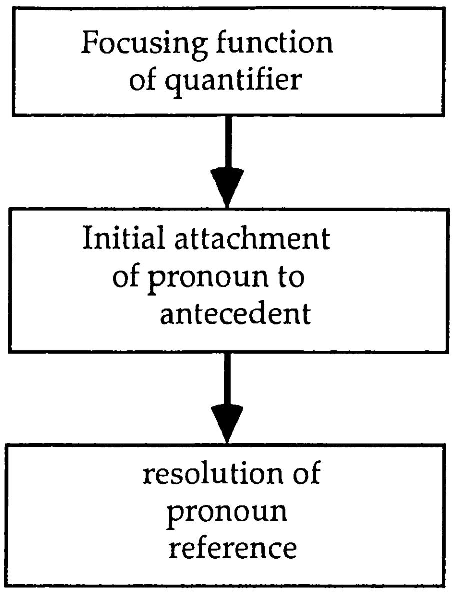


Figure 1: Proposed component processes of pronominal reference to a quantified NP

This heuristic is used when resolving pronouns that make reference to a quantified noun-phrase. For instance, when (2) is read as a continuation of (1), the experienced difficulty arises from an initial attempt to interpret the anaphor as referring to the reference set of the quantified noun-phrase, i.e. the set of MPs who attended the meeting. However, this appears anomalous, because the events described in (1) and those described in (2) cannot concern the same set of individuals. That is to say, the same set of MPs cannot both attend a meeting and go for lunch instead.

(1) A few of the MPs went to the meeting.
(2) They went for lunch instead.

According to my intuitions, this anomaly is unrecoverable, meaning that the reader cannot re-intepret the pronoun as referring to another set of individuals in order to reach a coherent interpretation of the pair of sentences.

Sentence pairs like (3) and (4) produce a temporary sense of anomaly when the antecedent sentence contains a monotone-decreasing quantified noun-phrase, and the second sentence describes an action or event that is congruous with the reference set. For example, (4) describes an event that is congruent with the reference set of MPs who did attend the meeting, an incongruent with the complement set of MPs who did not attend the meeting.

(3) A few of the MPs went to the meeting.

(4) They discussed funding levels for the NHS.

Both the reference and complement sets will reside in Explicit focus following monotone-decreasing quantification. However, the complement set will be the more accessible (i.e. focused) of the two. Therefore readers will initially interpret the plural pronoun as referring to the complement set. This will results in an anomaly when the reader cannot integrate the events described in the two sentences. However, this is a temporary anomaly because the reader can easily re-interpret the pronoun as referring to the less accessible (i.e. comparatively unfocused) reference set.

This account is consistent with other observations. Sanford and Garrod (Sanford, Garrod, Lucas and Henderson, 1984; Sanford and Garrod, 1989) report what they call 'false-bonding' phenomena, where a pronoun appears to initially and anomalously attach to the focal antecedent before re-assignment to a more appropriate one. In (5) this results in an impression that the neutral pronoun *it* refers to Ireland, before it is understood to refer to the boat taken by Harry. This antecedent is represented in Implicit focus as the instrument of the verb *to sail.*

(5) Harry was sailing to Ireland. It sank without trace.

In this case, and my quantificational example, the anomaly is experienced during the second stage of processing, when the processor attempts to establish a match between the semantic properties of the anaphor and those of its potential antecedent. In both examples, it becomes clear at some point downstream of the pronoun that its description is incompatible with the initial attachment.

The co-referentiality of an anaphor and its antecedent can be evaluated in several ways. The most obvious of these is a comparison of gender and number information[11]. For instance, reference to a male character by the singular feminine pronoun *she* is clearly anomalous, while Garrod and Sanford (1985) report an unpublished study where reference to an individual by singular pronoun resulted in a longer reading time compared to less appropriate plural pronoun reference.

[11] Gender and number aggreement are usually viewed as syntactic (c.f. Garnham, 1985).

Beyond these factors, the processor looks for congruency between the discourse roles of the anaphor and its antecedent. The process of doing this is described as a **pragmatic inference**, because the reader must infer a relation between the two discourses on the basis of general or specific background knowledge about the described scenario. For example, in (6) the discourse role filled by the pronoun finds a match with the actor (John), because needing to trust someone is consonant with giving a confidence. However, in (7) the pronoun is interpreted as reference to the patient, because being trustworthy is a better reason for receiving a confidence that giving one.

(6) John confided in Bill because he needed someone to trust.
(7) John confided in Bill because he was trustworthy.

Such inferences clearly depend on detailed knowledge about the described situation - that people who confide are looking for someone to trust, and that being trustworthy is a desired characteristic of a confidante - but can also be influenced by other sentence level factors. For instance, Caramazza, Grober and Garvey (1977) observe that the anaphoric reference in (6) is more acceptable than that in (7), and attribute this asymmetry to a lexically-encoded verb bias. Verbs like *confided in*, they argue, have an 'implicit causality' where the subject of the verb is charged with causing the current event. Following conjunction by the causal connective *because*, the subject NP is preferred as the antecedent of pronominal reference from the subordinate clause. This means that John is the preferred antecedent in both (6) and (7). Other verbs like *blame* differ in their implicit causality and favour reference to the object NP, which explains the greater acceptability of (8) compared to (9). Caramazza et al found that subjects took longer to assign an antecedent when the pronoun referred to the disfavoured NP.

(8) John blamed Bill because he spilt the coffee.
(9) John blamed Bill because he had to find a scapegoat.

Yet, as Ehrlich (1980) notes, implicit causality cannot fully explain the referential process, because the anaphor is eventually attached to the appropriate antecedent even when this is opposed by the verb bias. Ehrlich argues that readers must use knowledge about the described situation to resolve the anaphor in these cases, and do so even in those cases where verb bias appears to aid the referential outcome (6 & 8). Instead of viewing implicit causality solely as a lexically-encoded property, she suggests that verbs help structure a model of the described events, and that other lexical items will interact with these constraints. For instance, Ehrlich demonstrates that manipulating the form of conjunction between the main and subordinate clauses changes the preferred pronoun antecedent. Substitution of *and* for *because* leads to no strong attachment preferences, while *but* reverses the preferences attributed to implicit causality, as in (10). This time John is the preferred referent.

(10) John blamed Bill but he spilt the coffee.

Later studies suggest a similarly complex interaction between gender and verb semantics. In one study Garnham and Oakhill (1985) reported a significant effect of gender, but no effect of verb causality. However, Vonk (1985) produced a contrary pattern of results where verb causality determined pronoun reference, and there was no effect for gender. In an effort to reconcile these contrary findings, Oakhill, Garnham and Cruttenden (1992) observed that matching the pronoun and referent role-descriptions was inferentially simpler for Vonk's materials than for the Garnham and Oakhill ones, and tested the hypothesis that gender cues predominate when complex pragmatic inferences are required to map the anaphor to a verb-role, and that the latter guides reference when the inferences are simpler. There was some evidence of this interaction, but stronger evidence that the two factors constrain different aspects of the inferential process. Verb congruity effects, it appeared, produce a role-to-role matching, while gender information cues a role-to-name matching. This is consistent with other claims for a separation of the mental representation of discourse roles and those entities which fill them (eg. Sanford and Garrod, 1981).

Given this model of reference processing, a good test of my claims about quantification would be to measure the ease of pronominal reference to the refset and compset partitions of a quantified sentence. Reference to the focal subset should be processed more easily, and more quickly, than reference to the unfocused one - because the latter case involves additional processing to revise an initial and anomalous attachment to the focal subset. Such an experiment will also be informative about the sort of inferences which are made during comprehension. This is a topic of current controversy. One side of it (McKoon and Ratcliff, 1992) claims that only a minimal number of inferences are made during reading, and specifically rule out instances of anaphoric reference which depend on situation-specific knowledge. For instance, McKoon and Ratcliff argue that only those inferences which are supported by well-known information from general knowledge, or information which is explicit in the text, are automatically computed. This suggests that anaphoric reference to the compset will not be processed during reading, because the compset is neither explicit in the text, nor represented as general knowledge. Instead it is represented as the complement of the described set, whatever that might mean in the given context. The following section takes some time to detail this debate about the extent of inference processing and type of inferences which are made during reading.

Text comprehension and inference processing

Early inference research (eg Dooling and Lachman, 1971; Bransford, Barclay and Franks, 1972; Barclay, 1973; Anderson and Pichert, 1978) argued that the outcome of text comprehension is a model of the described situation, which is cued by information in the sentence, and includes background knowledge about that situation, or information from other modalities (Bransford and Johnson, 1972). Other researchers (eg Bower, Black and Turner, 1979; Galambos and Rips, 1982; Sanford and Garrod, 1981; Anderson, Sanford and Garrod, 1983) suggest that the background knowledge used to construct these models has a modular organisation, and that sentential information can trigger the instantiation of a detailed representation for a given situation[12]. Sanford and Garrod, for instance, argued that situation-specific packets of background knowledge, which they call **scenarios**, fill the implicit focus partition described in their theory of text processing. Once a scenario has been triggered by sentential input, subsequent sentences are interpreted with respect to this structure, unless they include cues, such as temporal adverbial phrases, which instruct a change of scenario (Anderson et al, 1983; Fauconnier, 1985).

This was intended to contrast with an interpretative approach (Katz and Postal, 1965) which assumed that meaning resided solely in the content of the sentence, and can be derived by assigning semantics to a deep structure representation following syntactic analysis of the sentence. However, the interdependence of sentential content and situation-specific or general knowledge is easily demonstrated. For example, sentences (11) and (12) describe very different situations despite their surface and propositional similarity. In (12) the wicket-keeper is understood to have physically caught the ball, but in (11) the policeman signals the traffic to stop without any such physical contact.

---

[12] Minsky, (1975), Schank and Abelson (1977) and Sanford and Garrod (1981) provide a more theoretic treatment of modular knowledge organisation and its application to language understanding.

(11) The policeman held up his hand and stopped the traffic.
(12) The wicket-keeper held up his hand and stopped the ball.

The same argument applies to the interpretation of under-specified verbs (Garnham, 1979). For instance, the verb *cut* takes a different sense in (13) and (14).

(13) The gardener cut his friend's grass.
(14) The barber cut his friend's hair.

In one classic experiment, Bransford et al also showed that inferences are controlled by an integration of sentential content and background knowledge. Subjects who read sentences like either (15) or (16) were later tested on their recognition memory for a list of sentences including (17) and (18). Those who had read (15) did not recognise (17) or (18) as previously presented sentences; but those who read (16) falsely recognised (18), but not (15). Bransford et al concluded that the mental representation of (16) includes the inference that fish swimming underneath the turtles will also swim underneath the log. This is based on knowledge about the likely relationship between the objects and transitive properties of the spatial preposition *on*. However, the preposition *beside*, which is used in (15) does not produce the same model of events or license this inference.

(15) Three turtles rested beside a floating log and a fish swam beneath them.
(16) Three turtles rested on a floating log and a fish swam beneath them.
(17) Three turtles rested beside a floating log and a fish swam beneath it.
(18) Three turtles rested on a floating log and a fish swam beneath it.

These days it is beyond dispute that the final outcome of text comprehension is a mental representation which integrates sentence content and at least some background knowledge. Even those who favour minimal inferential activity (eg Kintsch, 1988; McKoon and Ratcliff, 1992) concede that such a model is ultimately represented, although they question the validity of models which depend on a lot of situation-specific knowledge. The debate has since moved on to differentiate between those inferences which take place during reading, and others which are delayed, or made in response to task demands.

The above Bransford et al study illustrates the distinction. It is not clear whether the inference observed there was made when reading the original sentence, during the period between reading the sentence and the recognition test, or at the time of the recognition test. McKoon and Ratcliff (1991, 1992; also Perfetti, 1994) suggest that knowledge-based inferences like these will be made off-line because they are computationally demanding. Perfetti claims that it is more efficient to make a limited number of easy inferences during reading, when there is competition for scarce processing resources, and wait until later to devote these resources to more computationally demanding ones, than to attempt to generate all manner of inferences during reading.

Yet the perceived complexity of inferential processes depends on how those inferences are defined. Some researchers, like McKoon and Ratcliff, define those inferences which are made during the reading of a sentence as operations on a propositional database[13], which is the output of prior syntactic analysis (see also Kintsch, 1974; Kintsch and van Dijk, 1978). They suggest that other, knowledge-based inferences are delayed until a discourse model is constructed from an integration of propositions and knowledge about the described situation. Other researchers (eg Sanford and Garrod, 1981; Johnson-Laird, 1983; Garrod, 1985; Garnham, 1987; Oakhill, Garnham and Vonk, 1992) consider that the bulk of inferential activity takes place at the level of the discourse model, and differ over the need for propositions. Although Johnson-Laird, Garnham and their associates include an intermediate surface representation of text, they suggest that only the most limited inferential activity occurs at this level, such as some forms of elliptical reference (cf. Sag and Hankamer, 1984; Carreiras, Garnham and Oakhill, 1993). Sanford and Garrod (1981; Garrod, 1985; also Sanford and Moxey, 1995) eschew propositions for a direct mapping of text onto a model of the described situation. Other, less partisan definitions of inference, such as the creation of new semantic information from old (Rickheit and Ströhner, 1985), only finesse the dispute between propositional and model-based theories.

According to propositional theories, the major inferential task is to produce connectivity between propositions, either by argument overlap, co-reference (including anaphoric reference), or via the explicit use of connectives. Because these inferential processes take place during reading and draw upon limited STM (Short-term memory) resources, they only connect propositions which are in close proximity, no more than two sentences apart, and are consequently described as local coherence inferences. More global inferences, such as those

---

[13] These propositions are intended to approximate those of propositional calculus. However, research in logical semantics (cf Johnson-Laird, 1982) has established that propositional calculus, and even first order predicate calculus, are incapable of representing the complexity of natural language.

needed to achieve co-reference or causality between distant propositions, are ruled out because they operate on more propositions than can be held in STM. Knowledge-based inferences are also ruled out because the processor would need to orchestrate an even greater amount of information. McKoon and Ratcliff stress that only those inferences that are necessary for the reader to comprehend a piece of text are made during reading. McKoon and Ratcliff argue that these are usually inferences that produce local connectivity in the text and require minimal general knowledge. For example, general semantic knowledge will support a co-referential interpretation of *the dog* and *the collie*, or *the burglar* and *the criminal* (McKoon and Ratcliff, 1981).

Some studies of anaphoric reference are cited as evidence of this local coherence constraint. Clark and Sengul (1979), for instance, report reading time studies where anaphoric reference by both pronoun and definite NP to an antecedent in the immediately preceding clause was markedly easier than when the antecedent was two or three clauses back in the text (see 19, where the three potential antecedents are italicised).

(19) *A broadloom rug* in rose and purple colours covered the floor.
Dim light from *a small brass lamp* cast shadows on the walls.
In one corner of the room was *an upholstered chair*.

They surmised that the immediately prior clause takes privileged status in working memory, while McKoon and Ratcliff interpret this as evidence that only the propositions which form that clause are still resident in STM and open to reference from the currently processed clause. Other evidence comes from eye movement data reported by Ehrlich and Rayner (1983), which also suggests that reference to nearby antecedents is easier than reference to more distant ones. They observed an increased reading time following pronominal reference when the antecedent was distant in the text.[14]

However, a focus based account of anaphoric reference, like the one described in the previous chapter (also Sanford and Garrod, 1981; Garrod and Sanford, 1991) also predicts that recently mentioned NPs are the most accessible to anaphoric reference, unless textual cues have been used to focus on more distant ones. This means that neither Clark and Sengul, nor Ehrlich and Rayner can separate minimalist and model-based theories. The minimalist position, in

[14] The relationship between eye-movement data and inferential processing is discussed in chapter 5. Generally it is assumed that there is a sufficient coupling between eye-movements and text processing to attribute increased gaze duration to increased processing (cf Pollatsek and Rayner, 1991).

contrast, is vulnerable to examples of anaphoric reference based on situation-specific knowledge. These include the studies described in the previous section (Caramazza et al, 1977; Ehrlich, 1980; Oakhill and Garnham, 1985; Vonk 1985).

Other experimental demonstrations of the facilitation of pronoun resolution by situation-specific knowledge are reported by Sanford, Garrod and their colleagues. In one study, Anderson, Sanford and Garrod (1983) describe reference which depends on the continued accessibility of a cued scenario (a packet of situation-specific knowledge). They found that temporal adverbial phrases which expressed a time-shift beyond the current scenario reduced the availability of scenario-bound antecedents. For instance, in (20) the main character is introduced by name, and a secondary character is described by discourse role. Anderson et al manipulated the temporal adverbial phrase of the penultimate sentence to shift time either within or beyond a time constraint associated with the scenario (in this case the length of a film). Subsequent anaphoric reference (by feminine personal pronoun) to the main character was unaffected by this time-shift[15], but reference to the scenario-bound character (this time by masculine personal pronoun) took longer following a time-shift out of the scenario.

(20) <u>At the cinema.</u>
Jenny found the film rather boring.
The projectionist had to keep changing reels.
It was supposed to be a silent classic.
[Ten minutes / seven hours later] the film was forgotten.
[He / she] was fast asleep.

In a subsequent study, Garrod and Sanford (1985)[16] used a measure of spelling error detection latencies to demonstrate the detection of referential anomalies during reading, where this depended on situation-specific knowledge. Subjects read a number of passages like (21) where the pronoun could be resolved on the basis of gender but was followed by a verb which was either consistent or inconsistent with the discourse role of that antecedent. In this example, it is more likely that Elizabeth will *sink* than *jump* in the example material. Conversely, the lifeguard is more likely to *jump* than *sink*. When these verbs

---

[15] However, in continuation studies, Anderson et al noticed a tendency to use proper name or full noun-phrase to refer to the main character. Vonk, Hustinx and Simons (1992) also observed that reference by proper name can be used to signal a shift in topic.
[16] An eye-tracking study conducted by Garrod, Freudenthal and Boyle (1993) replicated these findings and is reviewed in chapter 5. A description of the earlier study is sufficient for present purposes.

were misspelled, Garrod and Sanford found a shorter error detection latency when they were consistent with discourse roles of the pronominal antecedents. This indicated either that interpretation of the predictable verb was facilitated by situation-specific knowledge, or that the unpredictable verb was registered as anomalous.

(21) Elizabeth was a very experienced swimmer and wouldn't have gone into the pool if the lifeguard hadn't been nearby. But as soon as she was out of her depth she started to panic and wave her hands about in a frenzy. Within seconds [she / he] [*jumped / sank*] . . .

A final observation by Garnham (1993) makes a nonsense of McKoon and Ratcliff's claims. He notes that their prohibition of knowledge-based inferences forces McKoon and Ratcliff to conclude that knowledge about the likely instrument in (22) is unavailable as this sentence is read, yet once (23) is used as a continuation, the same knowledge is needed to make the local coherence inference that the spoon was used to stir the coffee.

(22) Mary stirred her coffee.

(23) The spoon was dirty.

Taxonomies of inferences during reading

Theories of inference processing during text comprehension are successful to the extent that they produce a principled account of which inferences are made as the sentence is read and which are delayed or made in response to task demands. Two extreme positions can be ruled out immediately. It is self-evident that some inferences are made during reading. I have, for instance, already reviewed some evidence of anaphoric resolution during reading. Similarly, it is also self-evident that not all possible inferences are drawn (if only because there are infinitely many possible inferences). These include problem-solving inferences found in riddles and some forms of logical reasoning. Consider, as an example, the following syllogistic premises (24 & 25). Johnson-Laird (1983) describes this form of syllogism as particularly difficult, and reports a study where no subject inferred the correct conclusion (26). This is an inference which does not appear to be drawn either during reading or even when consciously evaluated. Not that this means that all logical inferences are as difficult. Johnson-Laird also gives an example of an easy syllogistic inference embedded in discourse, while Lea, O'Brien, Fisch, Noveck and Braine (1990) report evidence of propositional reasoning during normal reading.

(24) All of the bankers are athletes.
(25) None of the councillors are bankers.
(26) Some of the athletes are not councillors.
(from Johnson-Laird, 1983)

Most theories of text comprehension take a stance somewhere between these extremes. I have already criticised one theory (McKoon and Ratcliff, 1992) which argues for minimal inferences during reading. They, in turn, have attacked the mental models framework as a description of unlimited inference-making (see eg Garnham, 1992; Glenberg and Mathew, 1992; Zwaan and Graesser, 1992; Carreiras, 1993; for responses to McKoon and Ratcliff). Given my earlier description of mental model representations as only partial semantic representations, and the claim that communication is an attempt to maintain shared partial models (also Hall-Partee, 1979), it seems clear that McKoon and Ratcliff are attacking a straw man account of mental models. Nevertheless, mental models theorists, like those from more minimalist perspectives, have so far failed to specify which inferences are made during reading. The remainder of this chapter surveys several failed attempts at such a specification, and a final one which may be more promising.

One strategy has been to build a taxonomy of inferences where a line can be drawn between those which are made during reading and others which are made later or in response to task demands. The most obvious scheme classifies inferences according to the type of information they add to the representation. One category is composed of anaphoric and co-referential inferences, which establish links between different descriptions of the same token. Another, similar category instantiates general terms (eg infer *cow* from *milk the animal*), while causal inferences link events (eg Keenan, Baillet and Brown, 1984), and instrumental inferences instantiate instrument roles for verbs (eg infer *hammer* from *strike the nail*). Most contentiously, there are predictive inferences which infer probable subsequent events. For example, McKoon and Ratcliff (1986) argue that if predictive inferences are made during the reading of the sentence, then this should lead subjects to conclude after reading sentence (27) that the actress died.

(27) The director and cameraman were ready to shoot closeups when suddenly the actress fell from the 14th storey.

Less well-researched categories include inferences about characters' emotional states (Gernsbacher and Robertson, 1992), and thematic and goal-based inferences (eg Long, Golding and Graesser, 1992; Long and Golding, 1993).

Although it is generally concluded that anaphoric and co-referential inferences are made during reading (eg Garnham, 1982; Sanford and Garrod, 1989; McKoon and Ratcliff, 1992; Graesser and Kreuz, 1993), there is mixed evidence of the processing of other inference types (cf. Keenan et al, 1991, for a collation of studies on these types of inference). As an illustration, Keenan et al list two demonstrations of instrumental inference during reading (Paris and Lindauer, 1976; McKoon and Ratcliff, 1980), and three experiments where they failed to occur (Singer, 1979; Dosher and Corbett, 1982; Lucas et al, 1987). These studies used a variety of experimental techniques, including a self-paced reading study by Singer. In another instrumental inference study not listed by Keenan et al, Garrod and Sanford (1982) produced positive results using the same self-paced reading methodology. Cotter (1984) subsequently replicated both the Garrod and Sanford, and Singer results. She concluded that Garrod and Sanford found evidence of instrumental evidence because their instruments were integral to the meaning of the verbs used. For example, *a key* is part of the meaning of the verb *to unlock*. However, the instruments and verbs used by Singer were not so closely tied. This suggests that a categorisation of inference by the type of information which is added to the mental representation is not sufficient to determine whether an inference will occur as the sentence is read.

Other attempts to categorise inferences do not draw distinctions based on the type of information that is inferred, but by the function of the inference and the required knowledge resources. In particular, Carpenter and Just (1977) distinguish between forward inferences which add information to embellish the representation, and backwards inferences which search for this information as it is needed to fill a gap. They argue that only backwards inferences are performed during reading, and in response to current processing problems. Backwards inferences include the above categories of anaphoric and causal inferences, while instrumental and predictive inferences are interpreted as examples of forward inferences. Keenan et al (1991) describe backwards and forwards inferences as bridging (see also Haviland and Clark, 1974) and elaborative inferences respectively, to capture the supposed superfluousness of forward inferences.

Van den Broek, Fletcher and Risden (1993) further subdivide the backwards inference category into connecting inferences which link the currently processed statement to recently processed text, which may still reside in STM; reinstatements which connect the current statement to more distant text which is now in long-term memory (LTM); and backwards elaborations which draw on the reader's general knowledge. Because backwards inferences constitute a search process in response to current processing difficulty, it is possible that they will fail on some occasions (cf. Bosch, 1988, for examples of personal pronouns without antecedents). Van den Broek et al's sub-categorisation predicts that inferences about local aspects of the text are most likely to be successful. However, the distinction still fails to explain why forward inferences are made during reading under some conditions, but not others.

The minimalist position draws a different functional distinction, but results in similar claims. As described in the previous section, McKoon and Ratcliff (1992) distinguish between inferences which contribute to the local coherence of the representation, and others which contribute to the global coherence. Only the former, they argue, are automatically made during reading, and include anaphoric and co-referential inferences, and perhaps causal inferences. Inferences which draw on knowledge of the text stored in LTM, or general knowledge are not made automatically, but only when local inferences fail to maintain coherence, or as special strategy. For example, someone studying a text with the aim of producing a critical review may produce more knowledge-based inferences than someone browsing through a newspaper. McKoon and Ratcliff consider the mixed experimental evidence for forward inferences as evidence of their idiosyncratic nature. Perfetti argues (following Fodor, 1983) that such inferences are not amenable to scientific study because they lie outside the language-processing module, and are hard to quantify.

Apart from the weaknesses of the minimalist account which have already been noted, including the fact that anaphoric reference is often dependent on situation-based knowledge, other researchers (eg Garnham, 1992; Singer 1993) observe that the account is untestable as it stands, because it does not provide any independent means of determining which inferences are automatic, and which are strategic.

A third strategy emphasises extra-linguistic factors such as the motivation of the reader and nature of the text (Graesser and Kreuz, 1993; Graesser, Singer and Trabasso, 1994). Like McKoon and Ratcliff, Graesser and his colleagues suggest that the motivation of the reader will determine the degree of inference-making. More inferencing will take place during purposeful reading, when the reader has particular goals, or an interest in the text. Under other conditions, such as those which typical of the laboratory setting, readers will scan text at a shallower level and make fewer inferences. Inferences are also more likely to be made when the reader has sufficient background knowledge about the subject of the text (cf. Vonk and Noordman, 1991, discussed in the following section), and when the text promotes inferential activity. Although Graesser et al do not say what it is about a text which promotes inferences, they assert that the poor quality of most experimental texts tends to mitigate against a high degree of inferential activity.

Graesser et al propose a hierarchy of inferences where those at the bottom, including referential and causal inferences are most likely to be made, even under degenerate reading conditions, progressing upwards to include thematic and emotional inferences which require higher motivation by the reader and well-constructed texts. However there does not appear to be any principled reason for this hierarchy, other than that those inference types at the bottom of the scale are reliably observed in experimental studies, and those towards the top are only rarely observed. This still begs the question of why some inferences appear more easily made than others.

Semantic and pragmatic constraints on inference-making

A final classification scheme proposed by Vonk and Noordman (1991) is more appealing. They classify inferences on two dimensions. One of these is the distinction between necessary and elaborative inferences already mentioned. Their other dimension is more novel, and interesting. It distinguishes between what Vonk and Noordman call logically necessary and possible inferences[17]. Those in the logical necessary category include entailment, presupposition, conventional implication and transitive inference.

Although logical necessity is intended to describe the deductibility of inferences from a propositional description of the text (Vonk and Noordman, 1991, pp453), presumably by the application of logical inference rules, the above sub-categories describe semantic properties of sentences, which allows a narrower interpretation of logically necessary inferences as just those inferences which are under semantic control. For example, Kempson (1977) defines entailment as the necessary truth of a second sentence given the truth of the first. So the truth of (28) necessitates the truth of (29), because the property of being a bachelor necessarily implies the property of being male (and, strictly speaking, of marriageable age).

(28) That person is a bachelor.
(29) That person is a man.

A presupposition describes another truth relation between sentences. When reading a sentence like (30), one presupposes the existence of a king of France in order to evaluate the truth of that sentence. Entailment and presupposition differ in that the falseness of (28) precludes a truth value for (29), while a non-existent king of France means that the truth of (30) is indeterminate.

(30) The King of France is bald.

Conventional presupposition is a non-truth-conditional property of some lexical items (Grice, 1961, in Levinson, 1982). Grice argues that although the conjunction *but* has the same truth table as *and*, it is used differently, to contrast the connection between its conjuncts (see also Halliday and Hasan, 1976). Finally, transitive inference is a property of other function words (eg *behind, in front of, to the left of, to the right of;* cf. Johnson-Laird, 1983) whereby a transitive

---

[17] Vonk and Noordman also refer to logically necessary and possible inferences as invited and pragmantic inferences respectively. Although they use the former pair of descriptions more widely, the latter pair are more appropriate. I hope to show that they capture a distinction between the semantic and pragmatic control of inference.

relationship of the form *A is behind B* and *B is behind C*, implies that *A is behind C*.[18]

In contrast, possible inferences are expectations based on general or situation-specific knowledge, given the described state of affairs. Following the pragmatics literature (Levinson, 1982), these are more appropriately described as implicatures and are defeasible, unlike semantic inferences (conventional implicature is also non-defeasible: Levinson, 1982). This means, as Vonk and Noordman demonstrate, that implicatures can be negated by subsequent text, but the negation of a semantics-based inference produces a contradiction. One example of a possible inference is to conclude from (31) that John broke the pitcher. Subsequent text can deny this inference (eg 32) without contradiction.

(31) John slipped on a wet spot. He dropped the delicate pitcher on the floor.
(Johnson, Bransford and Solomon, 1973)

(32) It was undamaged because of the thick pile carpet.

The defeasibility of possible inferences suggests that these are weaker than semantics-based ones. It should be no surprise then, to find that there is inconsistant evidence for elaborative inferences during reading. Johnson et al (1973) showed, using an off-line sentence-recognition task, that subjects who read (31) inferred that the pitcher broke; but subsequent studies suggest that such inferences are not routinely made during comprehension. Neither Singer and Ferreira (1983), nor Potts, Keenan and Golding (1988) found any evidence of predictive inferences during reading, while McKoon and Ratcliff (1986) report equivocal results across a range of measures. Predictive inferences were observed on tests which were administered after reading (five minutes after the reading phase), or when subjects could take time to respond. On more immediate tests, or tests with a response deadline, the inferences were less apparent, or not observed at all. More recently, O'Brien et al (O'Brien, Shank, Myers and Rayner, 1988; Garrod, O'Brien, Morris and Rayner, 1990) reported the occurance of predictive inferences during reading when the context sufficiently constrained their likelihood.

McKoon and Ratcliff (1986, 1990, 1992) argue that forward inferences are only 'minimally-encoded' during reading. This emphasises that inference-making is

---

[18] Although it must be noted that spatial relation terms do not always give rise to transitive inference. For example, given that *B is to the right of A* and *C is to the right of B* it is not necessarily true that C is to the right of A. A, B and C may form of circle, for instance.

not an all-or-none affair, but that inferences will vary in their strength of encoding and specificity. If an inference is only partially encoded then subjects will show longer response times on tests of this inference compared to fully encoded ones, but less time than needed to generate the full inference at the time of test. McKoon and Ratcliff used a deadline recognition test to test for predictive inferences after subjects had read sentences like (27), repeated below.

(27) The director and cameraman were ready to shoot closeups when suddenly the actress fell from the 14th storey.

They expected that subjects would infer that the actress died, and asked them to indicate if the word *dead* had appeared in the sentence. Such an inference should slow decision time and increase the false recognition rate relative to a control condition. Further, because subjects were trained to respond within 650msec of the test word presentation, there was insufficient time to generate the inference in response to the probe. However, McKoon and Ratcliff observed only a small increase in error rate, and concluded that the inference had been made, but not strongly encoded.

At the same time, non-specificity means that there may be a number of competing possible inferences. For example, a reader might infer after reading (27) that the actress died, was injured and survived, or was miraculously unhurt. Now consider what this means for (32). That the pitcher broke is but one possible inference. It is also possible that it did not break, or that different aspects of the situation are salient - perhaps some consequence of the pitcher breaking (eg 33 and 34).

(33) No-one would taste the expensive Burgundy wine.
(34) John was grief-stricken. The pitcher was a gift from his recently deceased grandmother.

If the pitcher breaking belongs to a cluster of possible inferences it may be falsely recognised in an experimental paradigm like that used by Johnson et al, but not so evident on measures of inferential activity during reading. However, if the context conspired to promote this inference (following O'Brien et al, 1988; Garrod et al, 1990), or if there was a semantic cue to the outcome (*crashed* in 35, cf. Loftus and Palmer, 1972), then the inference that the pitcher broke is likely to be stronger or more specific, and be observed during the reading of the sentence.

(35) John slipped on a wet spot. The delicate pitcher crashed to the floor.

A separation of semantic and pragmatic-based inferences is also possible for other inference types. The first section of this chapter catalogued semantic and pragmatic constraints on the resolution of anaphoric reference. Semantic factors include gender constraints and the influence of verb and conjunction constraints, but an anaphor will also be resolved to the appropriate antecedent on the basis of knowledge about discourse roles and properties.

Vonk and Noordman (1991) find a similar interaction of semantics and pragmatics behind the generation of causal inferences. They identify *because* as a semantic indicator of a causal relation between two clauses (see also Halliday and Hasan, 1976), but demonstrate that the likelihood that a subject will make this inference is mediated by pragmatic knowledge about the conjunct clauses. In one experiment, subjects who read (34) as part of a text about spraycans and the environment failed to infer during comprehension that good propellants will not react with other materials, but made this inference in a subsequent verification task. Vonk and Noordman argue that this is because subjects lacked specialised background knowledge about propellants. In a later experiment, using similar materials, they demonstrated that subjects with specialised knowledge about the subject matter of the sentence do make such inferences during reading.

(36) Chlorine compounds are frequently used as propellants, because they do not react with any other substances.

It is likely that a similar interaction of semantics and pragmatics will determine the interpretation of quantifiers. There was some evidence of this in the introductory chapters. For example, Chapter one reported directional monotone inference patterns associated with quantifier semantics, but also evidence of context-dependency given the non-extensionality, and possible non-conservativity, of some quantifiers. Then, in chapter two, we saw that quantifiers like *few* and *not many* exert a strong semantic preference for compset focus, but the content of continuations to sentences quantified by *few* depended on the availability of knowledge-based inferences about the compset. When the quantified sentence described a violation of expectations about the situation there were more explanatory continuations than when the sentence matched expectations.

At the same time the quantifier *only a few* was seen to have a weaker focusing function and be more open to contextual influence. For example, it puts focus on the compset when in the context of other negative lexical items or if pragmatically supported. Moxey and Sanford found that subjects were more likely to produce continuations which described the compset when the quantified sentence describes a situation that violates norm-expectations. For example, given the two quantified sentences in (37) and (38), Moxey and Sanford found that subjects produced more compset-focused continuations for (37) than for (38).

(37) Only a few of the MPs were at the meeting. They . . .
(38) Only a few of the children ate their ice-cream. They . . .

They argued that this difference arises because (38) describes a situation in which a small set of children enjoy the ice-cream, and this violates the expectation that most children enjoy ice-cream. Subjects therefore provided explanations of why the other children (the compset) did not enjoy the ice-cream.

Two experiments are described in Chapter 8 which more directly test the claim that focus patterns associated with *only a few* are mediated by contextual or pragmatic knowledge. This study used sets of materials in which the quantified sentence described a situation which either matched or violated expectations. For instance, (38) is an example of a norm-matching sentence, and (37) is an example of a norm-violating sentence. This is because the refset for (37) would be expected to be small (for the UK, at least) because it is unusual for women to be over six foot tall. Correspondingly, it is unusual for them to be under five foot tall, and in (38) a small refset violates this norm. If the quantifier *only a few* is used to mark the refset is small relative to expectations, and focus is mediated by pragmatics, then this predicts that focus is placed on the refset in (37), and the compset in (38).

(37) Only a few of the women were over six foot tall.
(38) Only a few of the women were over five foot tall.

The studies test this claim in two ways. In one experiment subjects are asked to provide completions to sentences which begin with a plural pronoun that refers to the quantified sentence. Following Moxey and Sanford (1987), it was predicted that the content of these completion sentence would describe the property of the focused subset. That is, the continuation sentence for (37) would describe the refset of tall women, but the continuation sentence for (38) would describe the compset of short women. The second study tested whether this same interaction was observed during comprehension, and measured subjects' eye-movements as they read sentences which referred to either a property of either the refset or compset partition of a sentence like (39) or (38).

Conclusions

This chapter began by describing a proposed processing model for pronominal reference to a quantified sentence. This claimed that a pronoun is initially interpreted as reference to the focused subset of the quantified sentence, but that this interpretation can be revised if it proves to be anomalous. That is, a pronoun is initially interpreted as reference to the refset of a monotone-increasing quantified sentence. In contrast, when the sentence contains a monotone-decreasing quantifier, the pronoun is interpreted as reference to the compset. According to the processing model, reference will appear anomalous when there is a mismatch between the description of the anaphor and that of the antecedent. The process of detecting this mismatch will depend on pragmatic inference.

The model predicts that such a mismatch will trigger processing to re-analyse the pronoun as reference to an unfocused set. This means that there will be a longer processing latency for reference to the unfocused subset of a quantified sentence. It also predicts an asymmetry for reference to monotone-increasing and -decreasing quantified sentences. According to the production data reviewed in the previous chapter, only the refset is represented following monotone-increasing quantification, but both the refset and compset are represented, and plausible antecedents following monotone-decreasing quantification. This suggests that anomalous reference can be more easily, and therefore more quickly, revised following monotone-decreasing quantification.

The remainder of the chapter reviewed a number of studies on the processing of anaphora and inference during reading in support of the predictions made by the model. In particular, there were a number of studies which supported the claim that pronominal reference is processed during normal reading, and that difficulties experienced in processing pronominal reference are reflected in the time it takes to read the anaphoric sentence. Moreover, there were some evidence (e.g. Garrod and Sanford, 1985) that both focus and pragmatic inference constrain the resolution of pronominal reference during reading, and that readers detect a mismatch between the description of the anaphor and the focused antecedent.

The review of the inference processing literature also suggested that there are separable semantic and pragmatic constraints on the likelihood that inferences are made during reading. These factors may constrain the interpretation of quantified sentences in different ways. In particular, semantics appears to be the primary determinant of set focus in monotone-increasing and -decreasing quantified sentences. However, given a non-monotonic quantifier like *only a few*, there may be both semantic and pragmatic determinants of focus. The chapter reported some evidence for this, and outlined some further experiments which are reported in Chapter 8.

Chapter four:

The processing of anaphoric reference to a quantified noun-phrase

Introduction

This chapter reports the results of a pilot study and two self-paced reading (SPR) investigations of anaphoric reference to a quantified noun-phrase. The pilot study was used to generate a set of materials for the self-paced reading studies, and also for the two eye movement studies reported in the next chapter. The SPR studies test the three-stage model of anaphoric reference outlined in the previous chapter, and the claim that quantifiers like *a few* & *many* and *few* & *not many* are used to focus on different partitions of a quantified noun-phrase.

Chapter three's model of anaphoric reference proposed that some quantifiers divide noun-phrases into refset and compset partitions to place one of these partitions in discourse focus. Subsequent anaphoric reference is automatically attached to this focal subset, and the attachment is evaluated to either resolve the anaphor or force re-assignment. However, if both the refset and compset are both focused following monotone-decreasing quantification, then reference to these sets may be evaluated in parallel. Resolution is only possible if a semantic match is established between the anaphor and its antecedent. The model predicts both that reference to the focal subset is most felicitous, and that reference to the unfocused subset will require extra processing, which will be evident as a longer reading time for the anaphoric sentence (cf. Haviland and Clark, 1974). It also predicts that an anomalous temporary attachment of the anaphor and focal subset will produce the 'false-bonding' effect described by Sanford and Garrod (1989).

At the same time, I have described how quantifiers like *a few* and *many*, which have positive and monotone-increasing linguistic properties, appear to focus on the refset partition of an NP, while other quantifiers, like *few* and *not many*, which are negative and monotone-decreasing, tend to focus on the compset. Such preferences have been demonstrated in off-line production (Moxey and Sanford, 1987), but whether they occur during comprehension remains an open question. If they do, the above model predicts an increased reading time for reference to the compset following quantification by *a few* or *many*, and for refset reference following quantification by *few* or *not many.*

However, this may be complicated by an asymmetry between the two types of quantifier. Monotone-increasing ones not only focus on the refset, but block reference to the compset. In contrast, monotone-decreasing quantifiers permit reference to both sets. This may be because the refset is a necessary representation, and always available for reference, while the compset is only a possible representation which is instantiated following monotone-decreasing quantification. The asymmetry predicts that reference to the unfocused compset may be more difficult following quantification by *a few* or *many* than reference to the unfocused refset after quantification by *few* or *not many.*

Both SPR studies used short three-sentence passages like (1) as experimental materials, where an anaphor described a property of either the refset (*their presence*) or compset (*their absence*) of a quantified noun-phrase. The remainder of the anaphoric sentence was designed to remain neutral between refset and compset interpretations.

The first experiment compared the reading times for these reference types following quantification by either the positive and monotone-increasing quantifier *a few*, or the implicit negative and monotone-decreasing *few*. The second experiment compared the positive and monotone-increasing *many* with the explicit negative and monotone-decreasing *not many.*

(1) A Public Meeting.
Local MP's were invited to take part in a public enquiry about proposals to build a nuclear power station. [Quantifier] of the MPs attended the meeting. Their [presence | absence] helped the meeting run more smoothly.

Pilot study:

An off-line measure of anaphoric reference to differentially focused subsets of a quantified noun-phrase.

Introduction

The original Moxey and Sanford (1987) study had subjects provide continuations to discourse fragments composed of a simple quantified sentence with a partitive construction, and a second sentence which began with the plural pronoun *they*. It was expected that subjects would interpret the plural pronoun as reference to the focal subset of the quantified noun-phrase, and that this preference would be reflected in the content of their continuations.

The present study used a forced-choice selection task in place of the sentence-continuation task. Subjects read a number of two-sentence passages, where the first sentence provided a context, and the second was a quantified sentence with the same construction as the Moxey and Sanford materials. Two possible continuation sentences followed each passage. Both began with the plural possessive pronoun *their*, and differed only in the noun possessed by it, which described a property of either the refset or compset. The sentences were all simple narrative continuations. This meant that they were more similar to completions that subjects gave to monotone-increasing quantified sentences than monotone-decreasing ones in Moxey and Sanford's (1987) sentence-continuation study. Completions to the monotone-decreasing quantified sentences tended to provide explanations of the small refset. One result of this is that the present sentences may be more felicitous continuations to monotone-increasing that monotone-decreasing quantified sentences. However, this was judged to be unavoidable, since explanatory continuation sentences would be even less felicitous following monotone-increasing quantification. According to the Moxey and Sanford data, there was only a tendency for explanatory continuations following monotone-decreasing quantification, and some simple narrative completions were produced under these conditions, but no continuations provided explanations of the small refset following monotone-increasing quantification.

In the present study, subjects were asked to indicate which of two continuation sentences best continued the narrative. It was again expected that subjects would prefer the one which referred to the focal subset: the refset following quantification by *a few* or *many*, and the compset following quantification by *few* or *not many*. An example material is illustrated in (2), with the critical noun italicised.

(2) A Public Meeting.
Local MP's were invited to take part in a public enquiry about proposals to build a nuclear power station. [A few / few] of the MP's attended the meeting.

1. Their *presence* helped the meeting run more smoothly.
2. Their *absence* helped the meeting run more smoothly.

The study was conducted in two phases. The first phase examined passages quantified by *a few* and *few*. Each passage was evaluated by ten subjects, and no subject saw any passage in both quantifier conditions. Passages were accepted as experimental materials if the anticipated response for both quantifier conditions was equal to, or exceeded, a 70% criterion (although, in practise this requirement was only necessary for the monotone-decreasing condition), and the passages were incrementally tested until 24 had passed the criterion. This generated a set of broadly equivalent materials. These 24 materials were then tested again for the quantifiers *many* and *not many*. No subject who saw the materials for *a few* & *few* did so for *many* & *not many*.

Results

The mean 'correct' continuations[19] for all four experimental conditions are illustrated in Figure 1 below. The final 24 materials from the first stage of the pilot produced a mean 8.58 (se=0.23) - from a maximum 10 - selected refset continuations when the passage was quantified using *a few*, and a mean 7.79 (se=0.20) when quantified by *few*. In the second stage, the same materials produced a mean 9.58 (se=0.16) refset continuations following quantification by *many*, and a mean 9.33 (se=0.18) compset continuations following quantification by *not many*.

[19] 'Correct' continuations are those which made the predicted type of reference. Continuations were expected to describe the refset following quantification by *a few* or *many*, and the compset following quantification by *few* and *not many*.

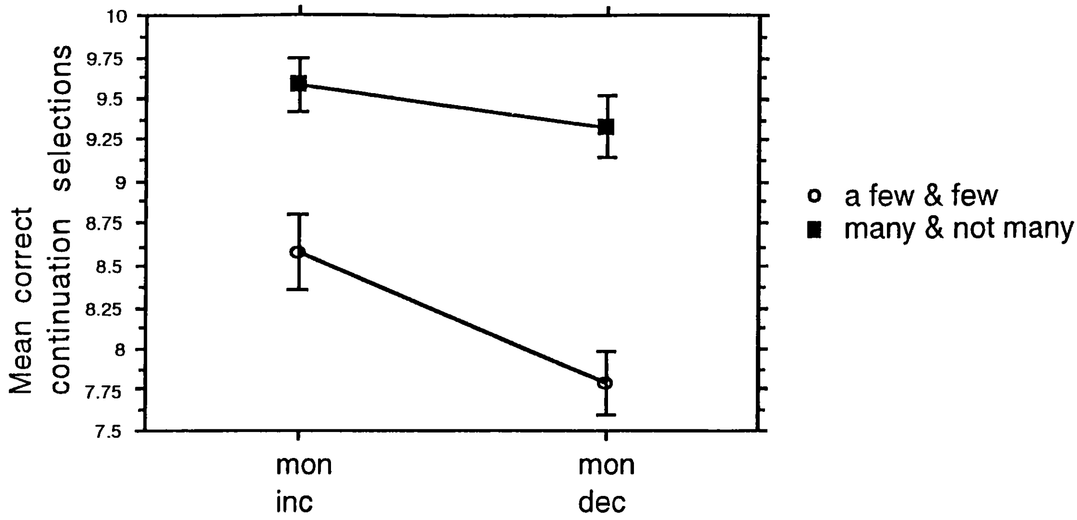


Figure 1: Mean 'correct' continuations (with standard error bars) for *a few*, *few*, *many* and *not many*

The overall results were analysed using a 2x2 related ANOVA - using a by items analysis only, because the first pilot did not use the same subjects to evaluate all of the materials - and produced a highly significant effect for the quantifier pair ($F(1,23)=57.55$, $p<0.001$), where there was a greater incidence of 'correctly' selected continuations sentences for *many* and *not many*. There was also a significant difference between the monotone-increasing and decreasing pairs ($F(1,23)=6.59$, $p<0.018$), with a lower incidence of 'correct' continuations following monotone-decreasing quantification. There was no evidence of an interaction ($F(1,23)=2.61$, $p<0.120$).

Discussion

The most pertinent observation is that across all of the 24 passages monotone-increasing and monotone-decreasing quantifiers exhibited the expected contrasting focusing function. Subjects chose those sentences which described the refset as the appropriate continuations for sentences quantified by *a few* or *many*, and showed a clear preference for sentences which described the compset as continuations to sentences quantified by *few* or *not many*.

There was also evidence of the anticipated asymmetry following monotone-increasing and monotone-decreasing quantification. I had suggested, based on the Moxey and Sanford (1987) off-line data, that there would be a more marked preference for continuations that described the reference set following quantification by *a few* and *many*, than for continuations that described the complement set following quantification by *few* and *not many*. This pattern was evident in the results. Subjects were more liked to select the sentence that described the reference set following quantification by *a few* than they were to select the sentence that described the complement set following quantification by *few*. Similarly, they were more likely to select the sentence that described the reference set following quantification by *many* than they were to select the sentence that described the complement set following quantification by *not many*.

What fine-grained differences there were between the investigated quantifiers were less expected. Subjects showed strong preferences for the 'correct' continuations following quantification by *many* and *not many*[20], but markedly weaker preferences following quantification by *a few* and *few*. This may indicate a difference in the strength of function of the individual quantifiers, with *many* a stronger predictor of refset focus than *a few*, and *not many* a correspondingly better predictor of compset focus than *few*. These differences may also appear during the reading of the sentence.

Experiment one:

A reading time analysis of anaphoric reference to refset and compset partitions of a quantified noun-phrase following quantification by *a few* and *few*.

Introduction

This experiment used the materials validated by the pilot study.
It was predicted that the referential preferences observed there would be reflected in measures of sentence reading time. Which means that there would be a longer reading time for anaphoric reference to the compset following quantification by *a few*, and for reference to the refset following quantification by *few*, compared to the other two referential conditions.

Reading times were also recorded for the context and quantified sentences, and although no reading time differences were expected for the context sentence across the experimental conditions, it was possible that there would be differences for the quantified sentence. This would arise from processing differences between monotone-increasing and monotone-decreasing quantification. According to my account, an extra processing stage is needed to

[20] In fact, the data suggests a ceiling effect for these responses.

derive the compset following monotone-decreasing quantification, because this remains unrepresented following monotone-increasing quantification. If this is true, then the extra processing stage may well be evident as an increased reading time.

Method

Design

There were two within-subjects manipulations: the form of quantification was crossed with the form of reference. That is, the subject NP of a quantified sentence was either quantified by the monotone-increasing *a few* or the monotone-decreasing *few*, and the following sentence began with an anaphor which either described a property of the refset or compset of the quantified sentence.

The cross-over design proved necessary because it is not clear what a control condition would look like for this experiment. The most obvious one is anaphoric reference to a bare NP. However, it is possible that the bare NP will naturally decompose into refset and compset partitions in some situations. Newstead (1994), for example, reports implicit quantification by verbs.

Subjects

Twenty-four subjects were used. All subjects were students attending an introductory summer school at the University of Glasgow, or undergraduate students there. Subjects were not paid for their participation.

Apparatus

Materials were presented on a 1024x768 pixel black and white monitor controlled by either a Macintosh SE30 or Macintosh IIsi computer running PsyScope experimental design software developed by Cohen, MacWhinney, Flatt, and Provost (1993) at Carnegie-Mellon University. A button box was used to measure subject responses. This had three buttons, coloured red, yellow and green from left to right. The central, yellow button was used to control the presentation of items and record reading time; the other two buttons recorded subject responses to comprehension questions. Red recorded YES responses and green recorded NO ones. The button box timer was accurate to one millisecond.

Materials

The experiment used the 24 materials validated by the pilot study. Each material had a title and was three sentences long. The passages were presented sentence-by-sentence and subjects were asked a comprehension question after every passage. This queried factual information about the first or second sentence in each passage, so as to avoid drawing attention to the referential manipulation in the third sentence. Half of the questions required a YES response, and the other half a NO response. Responses were recorded from the button box. An example material is given in (3).

(3) A PUBLIC MEETING
Local MPs were invited to take part in a public enquiry about proposals to build a nuclear power station.
Quantifier of the MPs attended the meeting.
Their **[presence / absence]** helped the meeting run more smoothly.
Was the enquiry about building a nuclear power station?

The materials were mixed with 18 filler passages (which also had comprehension questions), and subjects had a practice trial of 5 items before starting the experiment.

The experimental materials for each quantifier were divided into 4 blocks following a Latin Square design. Six subjects viewed each of the experimental blocks, and individual subjects viewed each material in only one of the experimental conditions. The materials were presented in a fixed order. The blocks with experimental materials, fillers and comprehension questions are listed in Appendix 1.

Procedure
Subjects read the following instructions before beginning the experiment:

"In this experiment you will read a number of short passages. The passages are presented a line at time, and a question will appear at the end of each passage.

Before each passage begins an asterisk (*) will appear on the screen. Look at the asterisk and wait for the title of the passage to appear. Once you have read the title press the YELLOW button for the first sentence of the passage. When you have read a sentence press the YELLOW button for the next one.

The question at the end of each passage will have a YES or a NO answer. Press the RED button if your answer is YES, and the

GREEN button if your answer is NO.

You should keep your fingers over the keys throughout the experiment and use the same finger for the same key all the time.

There will be a practice session first to make sure you know what to do.
Press the YELLOW button to start."

Subjects then read the passages sentence by sentence and at their own pace. A fixation spot appeared for in the top left corner of a presentation box prior to the presentation of each item. After 1500msec it was replaced by the passage title. Subjects then controlled the presentation of successive sentences using button-box responses. At the end of each item subjects were instructed to "Press the YELLOW button for next passage".

Results

8.2% of data points were lost prior to analysis because of recording errors with the button box.

Overall, subjects answered 9.6% of the comprehension questions incorrectly. Across the experimental conditions, this worked out as 11.1% errors for the *a few* / refset reference condition, 8.9% for the *a few* / compset reference condition, 7.8% for the *few* / refset reference condition, and 10.7% for the *few* / compset reference condition. A sign test compared the error rates for felicitous (*a few* / refset and *few* / compset) compared to anomalous reference (*a few* / compset and *few* / refset) condition, and found no significant difference ($x=3$, $N=10$, $p<0.18$).

The reading times for the context, quantified and anaphoric sentences were separately analysed using 2x2 within subjects ANOVAs.

An analysis of the mean reading times for the anaphoric sentence revealed a significant interaction of monotone quantifier and reference types ($F1(1,23)=36.22$, $p<0.001$; $F2(1,23)=10.34$, $p<0.001$). As can be seen from Figure 2, this is due to increased reading time for the two conditions making anomalous reference to the unfocused NP subset: that is, reference to the compset following quantification by *a few*, and reference to the refset following quantification by *few*. However, an analysis of the simple effect contrasts established that there was a highly reliable difference between the two reference conditions following quantification by *a few* (F1(1,23)=23.06, p<0.001; F2(1,23)=13.14, p=0.001), and the contrast between the two *few* conditions is only significant on a subjects analysis ((F1(1,23)=6.02, p<0.022; F2(1,23)=0.853, p=0.365).

Simple effects mean contrast also demonstrated that there were no reliable differences between the two felicitous reference, and the two anomalous reference conditions. A comparison of refset reference following quantification by *a few,* with compset reference following quantification by *few* was only significant by subjects (F1(1,23)=4.42, p<0.047; F2(1,23)=2.59, p<0.121); with no difference between compset reference following quantification by *a few,* and refset reference following quantification by *few* (F1(1,23)=0.06, p<0.810, F2(1,31)=1.19, p<0.287).

There were no significant main effects for the anaphoric sentence, and no significant effects were observed on analyses of the context and quantified sentences (all F's<1.4).

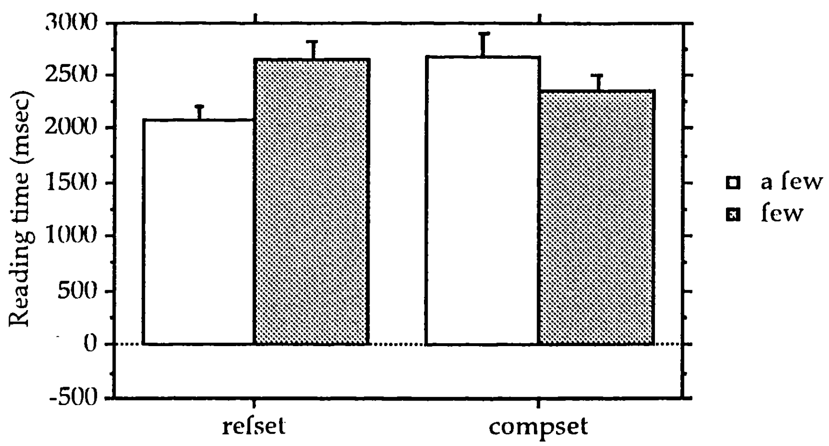


Figure 2: Mean reading time (with standard error bars) for sentences describing either the refset or compset partition of a noun-phrase quantified by *a few* or *few.*

Discussion

As predicted, there was a longer reading time for reference to the compset following quantification by *a few*. However, the difference in reading time for refset and compset reference following quantification by *few* was not so reliable. According to the subjects analysis, there was a longer reading time for refset reference. This was not supported by the items analysis, however.

Since previous studies have shown that sentence reading time can reflect difficulties in integrating a sentence with prior discourse (Haviland and Clark, 1974), and more particularly, the processing of anaphoric reference (Garrod and Sanford, 1977), it can be concluded that there is a greater processing difficulty when reference is made to the complement set of a sentence quantified by *a few*. This is consistent with an account where the monotone-increasing *a few* focuses on and enables reference to the refset, but blocks reference to the compset. In fact, the compset may even be unrepresented under these conditions. It was not possible to draw this or similar conclusions about the processing of reference following quantification by *few*. It appeared that reference was either equally easy, or equally difficult to process under these quantificational conditions. It suggests either that reference to an unfocused subset is more easily recovered from, or that both subsets are evaluated in parallel as potential antecedents, following quantification by *few*. The following experiment tested the same hypotheses for reference to sentences quantified by *many* and *not many*, to determine if it was possible to generalise the present conclusions to a broader range of monotone-increasing and decreasing quantifiers.

Experiment two:

A reading time analysis of anaphoric reference to refset and compset partitions of a quantified noun-phrase following quantification by *many* and *not many*.

Introduction

This second experiment replicated the previous study, but substituted the quantifiers *many* and *not many* for *a few* and *few*. It followed the same design and procedure and used the same materials; and what methodological differences there were are listed below.

Method

Subjects

The 24 subjects used for this experiment were drawn from the same subject pool as the previous experiment, but no subject was used for both. Subjects were not paid for their participation.

Materials

The experiment used the same materials and fillers as the previous study, but was run in conjunction with another discourse experiment which had 50 items. The order of the two experiments was counterbalanced across subjects.

Results

11.1% of data was lost prior to analysis because of errors associated with the button box.

An analysis of the question responses found 10.5% errors overall, with 9.5% in the *many* / refset reference condition, 12.5% in the *many* / compset reference condition, 10.0% in the *not many* / refset condition, and 10.2 % in the *not many* / compset reference condition. A sign test comparison of errors made in felicitous (*many* / refset and *not many* / compset) and anomalous (*many* / compset and *not many* / refset) reference conditions proved non-significant ($x=6$, $N=16$, $p<0.23$).

As before, the reading times for the context, quantified and anaphoric sentences were separately analysed using within subjects 2x2 ANOVAs.

The analysis of the anaphoric sentence produced a significant interaction of monotone quantifier and reference types ($F1(1,23)=22.21$, $p<0.001$; $F2(1,23)=25.93$, $p<0.001$), with increased reading times when the sentence described the unfocused noun-phrase subset (see Figure 3). Further analysis of simple effect contrasts again indicated that there was a highly reliable difference between the two reference conditions following quantification by *many* ($F1(1,23)=17.85$, $p<0.001$; $F2(1,23)=24.60$, $p<0.001$); and that there was a significant difference between the two monotone-decreasing conditions ($F1(1,23)=5.95$, $p=0.023$, $F2(1,23)=5.03$, $p=0.035$).

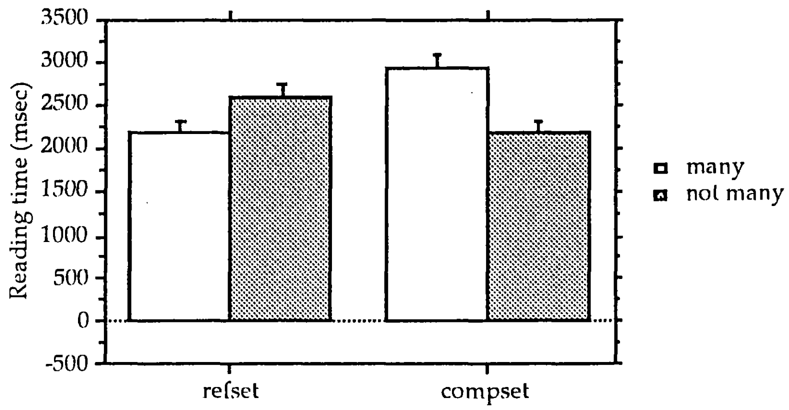


Figure 3: Mean reading time (with standard error bars) for sentences describing either the refset or compset partition of a noun-phrase quantified by *many* or *not many*.

Other simple effects means comparisons found no reliable difference between the two felicitous reference conditions, but a difference between the two anomalous reference ones. A comparison of refset reference following quantification by *many*, and compset reference following quantification by *not many* was non-significant (F1(1,23)=0.00, p<0.948; F2(1,23)=0.03, p<0.867). The comparison of compset reference following quantification by *many*, and refset reference following quantification by *not many* was marginal by subjects and significant by items (F1(1,23)=3.43, p<0.077; F2(1,23)=6.49, p<0.019); with a longer reading for anomalous reference following quantification by *many*.

There was also a main effect of quantifier type (F1(1,23)=5.02, p=0.035; F2(1,23)=3.54, p=0.073) which reflected an overall longer reading time for reference to sentences quantified by *many* than those quantified by *not many*. Finally, there was a main effect of reference type which was significant by items only (F1(1,23)=2.80, p=0.108, F2(1,23)=4.75, p=0.040), which reflected an overall longer reading time for reference to the compset of quantified sentence.

The analysis of reading times for the quantified sentence produced a main effect of quantifier type, which was significant by subjects only (F1(1,23)=4.53, p=0.044; F2(1,23)=2.38, p=0.136), and was due to a longer reading time for sentences quantified by *not many*. There were no other significant results (all F's <1.7).

Discussion

The results confirmed those for the previous study. There was a longer reading time for the anaphoric sentence when it referred to the unfocused subset of the quantified sentence. That is to say, there was longer reading time for the sentence when it referred to the complement set as compared to the reference set following quantification by *many*; and a longer reading time for the sentence when it referred to the reference set as compared to the complement set following quantification by *not many*.

The weak main effect for the quantified sentence (which was significant by subjects only) was in the predicted direction, with a longer reading time for monotone-decreasing quantified sentences. However, this may have been due to the extra word (*not*) in the monotone-decreasing condition.

Conclusions

The self-paced reading experiments produced the same pattern of effects following monotone-increasing quantification. That is, there was an increased reading time for the anaphoric sentence when it referred to a property of the compset following quantification by either *a few* or *many*. This allowed the conclusion that anaphoric reference to the compset is more difficult to process following monotone-increasing quantification, because this compset is either unfocused or unrepresented. Given the considerable difficulty that subjects have in reading the sentence under these conditions, it is very likely that the latter is true, and the compset is unrepresented.

There was a less reliable pattern of data following monotone-decreasing quantification, however. The results for *not many* were consistent with the experimental predications: there was a longer reading time for refset as compared to compset reference. This suggested that the compset is focused, and the preferred antecedent of anaphoric reference under this form of quantification. The same pattern of results was less reliable following quantification by *few*, however. It seems that both forms of reference are equally possible under these conditions. This is important because it indicates that these quantifiers do have a different function from monotone-increasing ones, but not in terms of contrasting focus. It may be more appropriate to view monotone-decreasing quantification as a means of introducing additional information into the mental representation, rather than as a means of reversing focal preferences.

While the experimental results permit conclusions about the ease in integrating the anaphoric sentence with prior discourse, and demonstrate that this depends on the form of quantification, the methodology was insufficiently fine-grained to enable an evaluation of the model of anaphoric processing which was outlined at the beginning of Chapter three. For instance, it is not possible to say whether the difficulties in integrating the anaphoric sentence with prior discourse are due to increased processing to recover from an anomalous interpretation of the anaphor as reference to the focused antecedent, or simply due to difficulties in recovering the unfocused antecedent from the discourse model. The former predicts that the pronoun is immediately interpreted as reference to the focused antecedent, and that this is detected as a referential anomaly at some point downstream, and probably at the following noun, when it describes a property of the unfocused subset. The experiments reported in the following chapter attempt to separate these possible accounts.

Chapter five:

The time course of plural anaphoric reference to a quantified noun-phrase.

Introduction

The experiments reported in this chapter are intended to provide a more detailed account of the processes involved in anaphoric reference to a quantified NP. Those reported in the previous chapter have established that some quantifiers are used during reading to focus on specific subsets of the quantified NP, and that the interpretation of subsequent pronominal reference is contingent on this function. They showed that monotone-increasing quantifiers like *a few* and *many* focus on and enable reference to the refset partition of a quantified NP, while blocking reference to the compset. In contrast, those which are monotone-decreasing, such as *few* or *not many*, serve to direct the representation of the compset, and permit pronominal reference to either subset. There was also evidence to suggest that the complement set is the preferred referent following quantification by *not many*. Yet the studies are largely uninformative about the locus and time course of these processes. For instance, it may be that the interpretation of a quantifier as an instruction to focus on a particular NP subset is made on reading that quantifier, or in response to anaphoric reference, or even deferred until the sentence is fully read. At the same time, it is not clear whether referential processes are triggered by the anaphor itself, or delayed until later.

A more fine-grained measure, such as a record of subjects' eye movements, may help answer these questions, and in so doing be informative about the component processes of anaphoric reference. The model described in Chapter three proposed that pronominal reference to a quantified NP is initially interpreted as reference to the focal subset, and if this interpretation proves to be wrong, is later revised to the unfocused one (or to a superset interpretation). If this is the case, then it should be possible to determine the point in processing where this initial assignment is registered as anomalous. This will be referred to as the locus of referential anomaly detection. We can infer that anaphoric processes were initiated at or before the locus of referential anomaly detection, and completed at or beyond this point.

Knowing the time course of processing can also help to further delineate those processes which are on-line and concurrent with reading, and others which are deferred until later. Some researchers (eg McKoon and Ratcliff, 1992; Perfetti, 1994) have argued that only the former are a necessary part of comprehension. The next section reviews some evidence of on-line anaphoric processing. Given that the referential processes described in this thesis are more complex that those which are usually the subject of psychological study, it will be interesting to see if these too are computed on-line. There may even be differences between reference under the two types of quantification.

The time course of anaphoric processing

While many of the anaphoric reference studies cited in Chapter 3 are just as uninformative about the time course and component processes of anaphoric references as those reported in the previous chapter, the data reported by some suggests that processing is at least triggered, and sometimes completed, on encountering the anaphor. In particular, they suggest that processing can be completed when sufficient cues to the referential outcome are localised at the anaphoric expression. For example, co-referential definite NPs like *the Prime Minister* and *John Major* (given a contemporaneous discourse setting) can in principle be resolved immediately (but see Garrod, Freudenthal and Boyle, 1994); and the same seems true of pronouns when gender or number are sufficient cues to their resolution.

The various paradigms used by the studies described in Chapter 3 may also be informative about different stages of anaphor resolution. For example, Sanford and Garrod (1989; also Garrod et al, 1994) argue that methodologies which interrupt the reading process with lexical decision or word recognition tasks are indicative of initiation processes. Such studies might describe the activation of candidate referents, and possibly the suppression of non-candidates, but do not indicate whether a commitment has been made to any particular antecedent. Other paradigms which present the same probes at the end of sentence (eg Corbett and Chang, 1983; Gernsbacher, 1989), or require subjects to indicate their referential decisions (eg Stevenson and Vitkovich, 1986), will describe the completion of anaphoric reference. Moreover, the task performed by subjects is also relevant. Cloitre and Bever (1988) had subjects read sentences containing either definite or pronominal reference to a previously-mentioned person or object. On encountering the pronoun subjects were presented with a target word that was either related or unrelated to the previously mentioned person or object. In one experiment, subjects had to decide if the target was a word or a non-word, and in another experiment they had to decide if it belonged a particular category (e.g. if it was a type of fruit). Cloitre and Bever found that subjects made faster lexical decisions word that were related as compared to unrelated to the antecedent when the sentence contained a definite referring expression, but no difference is decision latencies when the sentence contained a pronoun. In contrast, subjects made faster category decisions for related as compared to unrelated targets when the sentence contained either a definite referring expression or a pronoun. Cloitre and Bever concluded that the tasks were differentially sensitive to the processing of definite and pronominal reference.

Dell, McKoon and Ratcliff (1983), and O'Brien, Duffy and Myers (1986) used the former experimental paradigm to produce evidence of immediately initiated referential processes following definite NP reference. Both studies demonstrated that the recognition of a previously read NP is facilitated, relative to a control condition, by the immediately prior presentation of an anaphoric NP (Dell et al locate the effect within 250msec of presentation of the anaphor). For example, Dell et al conducted two experiments where subjects read passages like (1) completed by either the experimental sentence (2), or a control sentence (3).

(1) A burglar surveyed the garage set back from the street.
Several milk bottles were piled at the kerb.
The banker and her husband were on vacation.

(2) The$_1$ criminal$_2$ slipped$_3$ away$_4$ from the$_5$ streetlamp.$_6$
(3) A$_1$ cat$_2$ slipped$_3$ away$_4$ from the$_5$ streetlamp.$_6$

In one of the experiments subjects performed a word recognition task (at points 1 to 6 of the final sentence) on *burglar*, which had appeared earlier in the passage, and is a plausible antecedent of *the criminal*, but not of *a cat*. In the second experiment, subjects performed the same task on an associate word, *garage*, which was also presented earlier in the passage, and described by Dell et al as belonging to the same proposition as *burglar*. The results demonstrated that both the referent and associate were more quickly recognised as previously read words when tested at points 2, 3 and 4 of the experimental sentence than when tested at the same points in the control sentence. This suggests that *burglar* is registered as a candidate antecedent as soon as the anaphor is read and maintained as such until some distance downstream. It is also likely that it is selected as the appropriate antecedent at some point during this processing, but Dell et al cannot say when.

Speelman and Kirsner (1990) replicated the Dell et al results, using pronouns in place of anaphoric NPs, and concluded that pronominal reference is as rapidly initiated as definite NP reference. However, Greene, McKoon and Ratcliff (1992) disputed this conclusion, and used a speeded reading and recognition task to suggest that the recognition of a previously presented antecedent is more rapid following definite NP reference than following pronominal reference. Subjects read short passages like (4), word-by-word and at a fixed rate (each word was presented for 250msec, with the last word of each line presented for 300msec to allow for sentence wrap-up processing), while subjects had also to complete the recognition task within a fixed time period (1000msec). A recognition probe was presented at either points 1, 2 or 3 in the passage, testing recognition of the referent name (*Mary*, in this instance) relative to the non-referent name (*John*), or a non-referent control word (*dishes*).

(4) Mary and John were doing the dishes after dinner.
One of them was washing while the other dried.
Mary accidentally scratched John with a knife
and then$_1$ she dropped$_2$ it on the counter.$_3$

Although Greene et al found no evidence of facilitated recognition for the referent (*Mary*) compared to non-referent (*John*) conditions, these showed facilitated recognition relative to the non-referent control word (*dishes*) at all three test points. However, this only suggests that the named characters take privileged status relative to the non-referent control (see also Corbett and Chang, 1983), particularly as facilitated recognition was observed prior to pronominal reference (at point 1). When Greene et al replicated the earlier Dell et al study using the speeded reading and recognition task, they observed facilitated recognition of the referent noun-phrase (and its semantic associate), and concluded that their paradigm prevented time-consuming processing needed to effect pronominal reference, but allowed more rapidly completed and easily processed definite NP reference.

Other studies conducted by Corbett and Chang (1983), and Gernschbacher (1989), suggest subtler differences between NP and pronominal reference. For example, Corbett and Chang had subjects read sentences like (5), where the first mentioned character is referred to by proper name, pronoun or elliptical reference. The sentences were read word-by-word, but self-paced, and with a recognition probe presented at the end of each sentence. The response latencies indicated that recognition of the antecedent (*Karen*) was facilitated relative to the non-antecedent (*Emily*) whatever the reference type. However, an examination of response latencies for the non-antecedent probe (*Emily*) suggested slower recognition following proper name reference than following either pronominal or elliptical reference.

(5) Karen poured a drink for Emily and then [Karen / she /*ellipsis*] put the bottle down.

Corbett and Chang suggest that proper names activate the representation of potential antecedents, and suppress non-antecedents, while pronouns simply activate the representations of all plausible antecedents. Gernsbacher observed similar effects when she used a modified version of this experiment to examine the time course of facilitated antecedent recognition. She found that the appropriate antecedent of a proper name anaphor showed facilitated recognition (relative to a non-antecedent) immediately following the anaphor, but the same effect was weaker and delayed until the end of sentence for pronominal reference.

These results suggest the immediate initiation of proper name anaphora (just as Dell et al found for definite NPs) and delayed initiation of pronominal anaphora. They also suggest a more specific function for proper name anaphora. They activate the representation of potential antecedents and inhibit the representations of non-antecedents, while pronominal anaphora only seem to activate the representations of plausible antecedents. However, both of these studies used materials where pronoun gender failed to unambiguously select an antecedent, and it may be that pronominal reference is delayed and exhibits a non-specific function under these conditions. After all, no delay was observed in the Speelman and Kirsner study where gender did unambiguously select the appropriate referent.

Furthermore, other studies using a cross-model priming task to investigate anaphoric processing during speech comprehension suggest the immediate initiation of pronominal reference when gender is a useful cue. Shillcock (1982), for example, presented sentences like (6) to subjects, who made lexical decisions about target words presented at either points **a**, **b** or **c** in the text. In this example the targets were either *street* or *school*, where *school* is semantically associated with *the teacher*, a potential antecedent of pronominal reference. *Street* is a control word which is unassociated with either of the potential antecedents (the teacher or the train).

(6) The teacher **a** did not board the train, for the **b** simple reason that it/he **c** was not going to the South Coast of England.

Shillcock reported equal lexical decision latencies for the two words at point **a**, and a faster decision time for *school* at point **b**. At point **c**, however, lexical decision was faster for *school* following reference by masculine personal pronoun (*he*), but equivalent to the decision latency for *street* following reference by neutral pronoun (*it*). Shillcock concluded that the discourse representation of the teacher was inhibited when it proved an inappropriate antecedent (following neutral pronoun reference) but maintained when the reference was more appropriate.

Other studies reviewed by Nicol and Swinney (1989) demonstrate that potential antecedents of a pronoun are immediately re-activated unless these are ruled out by structural constraints, and suggested that semantic and pragmatic constraints select from this candidate set. In one experiment, subjects listened sentences like (7), where structural constraints mean that the reflexive pronoun (*himself*) can only refer to *the doctor,* and made a lexical decision to visual presentations of semantic associates of each of the potential antecedents (*the boxer, the skier, the doctor*). These were presented at the point in the sentence marked by an asterisk. The 'i' indices denote the syntactically preferred co-referential interpretation.

(7) The boxer told the skier that the doctor$_i$ for the team would blame himself$_i$ * for the recent injury.

Nicol and Swinney report that lexical decisions were significantly faster for a semantic associate of *doctor* as compared to a control word, but there was no difference between lexical decisions for semantic associates of *boxer* or *skier* compared to control words. Nicol and Swinney interpret this as evidence that syntax can be used to immediately constrain the candidate set of referents, and in this case select a unique referent, of a pronoun.

In a second study, subjects listened to similar sentences like (8), where the reflexive pronoun was replaced by an object pronoun (*him*). This time syntax only ruled out the doctor as an antecedent, but the boxer and the skier remained possible antecedents. Subjects again made lexical decisions to visual presentation of semantic associates of *boxer, skier* or *doctor.* These were presented at the point in the sentence denoted by an asterisk, and the syntactically acceptable co-referential interpretations are indicated by 'i' indices.

(8) The boxer$_i$ told the skier$_i$ that the doctor for the team would blame him * for the recent injury.

This time there was no difference in decision latencies for a semantic associate of *doctor* compared to a control word, but shorter latencies for lexical decisions to semantic associates of *boxer* or *skier* compared to control words. Nicol and Swinney concluded that syntax can constrain the candidate set of potential antecedents, and argue that semantics and pragmatics will then serve to select between these alternatives. This may be compatible with Shillcock's findings, since pronoun gender is often considered a syntactic constraint. Both sets of results suggest the immediate initiation of pronominal reference, but that the set of candidate antecedents are also immediately constrained by syntactic factors.

Later cross-model priming studies conducted by Tyler and Marslen-Wilson (1982) and Marslen-Wilson, Tyler and Koster (1993) were more informative about the locus of resolution, and examined discourse constraints on the resolution process. In the first of these studies (Tyler and Marslen-Wilson, 1982), subjects heard passages like (8) which were completed by one of the sentence fragments (10) to (12). They were then visually presented with a probe (either the pronoun *him* or *her*). Subjects were required to name this probe. Tyler and Marslen-Wilson reasoned that there would be a shorter naming latency (i.e. the time taken to begin to pronounce the probe) when the object pronoun was consistent with the context (*her* in all cases), and that this would indicate the prior resolution of proper names (*Philip*, in 10), pronouns (eg 11) and elliptical reference (eg 12). They reported such a difference, with a less marked effect for the elliptical condition[21].

(9) As Philip was walking back from the shop, he saw an old woman trip and fall flat on her face. She seemed to be unable to get up again.

(10) Philip ran towards . . .

(11) He ran towards . . .

(12) Running towards . . .

---

[21] This may have occured because elliptical reference is a low-level phenomena (see Sag and Hankamer, 1984) which may not attend to semantic information, but make an immediate and ballistic attachment to the focal referent. This may have been to the old woman in Tyler and Marslen-Wilson's study, which would produce a temporary anomaly and slow processing.

Marslen-Wilson, Tyler and Koster (1993) conducted a modified version of this experiment which demonstrated that gender cues, pragmatic inference and discourse focus are all used to immediately constrain the interpretation of a pronoun. Marslen-Wilson et al found that when gender disambiguated the pronoun this was the dominant constraint, and there was a shorter naming latency for the probe. However, the discourse focus of the antecedent and knowledge that the antecedent was an appropriate patient of the preceding verb-phrase (e.g. *ran towards* or *running towards* in 10, 11 & 12) also influenced the naming latency of the probe. There was a shorter naming latency for the pronoun probe when it referred to the antecedent which could be pragmatically inferred to be the appropriate patient of the verb. For instance, it can be pragmatically inferred that the old woman in (9) is the most appropriate patient of the verb-phrase *run towards* in (10) or (11), or *running towards* in (12). Finally, there was a shorter naming latency for the pronoun probe when the antecedent was in discourse focus. That is, when it was consistently referred to by the subject NP of each sentence in the discourse. Although all of these constraints appeared to influence the resolution process, Marslen-Wilson et al argue that gender was the dominant constraint, pragmatic inference a less strong constraint, while discourse focus was the weakest.

With the exception of the Greene et al (1992) studies, the evidence appears to favour the immediate initiation of referential processing on encountering either a definite NP, proper name or pronoun. Some of the data also suggests that definite NPs or proper names are resolved earlier than pronouns, although gender did not unambiguously select an antecedent in these studies. The speech processing experiments (Shillcock, 1980; Tyler and Marslen-Wilson, 1982) suggest that pronominal reference is rapidly resolved when gender is a useful cue. Vonk (1984) used measures of eye movements and fixation duration during reading to also demonstrate early pronoun resolution. She found a small but reliable increase in fixation duration on the pronoun when gender identified the antecedent, and increased reading time on the following verb when it was not a useful cue.

Another study by Stevenson and Vitkovich (1986) suggests that pragmatic information about verb roles is also rapidly used to determine the appropriate antecedent of a pronoun when gender was not a useful cue. Subjects read sentences like (13) to (15) and had to indicate which of the potential antecedents was referred to by an anaphor. The verbs in the second clause of the experimental sentence were either informative (*hurried* in 13 and 14) or uninformative (*forgot* in 13 and 14) about the referential outcome, while reference was made by either a personal pronoun (13 and 15) or ellipsis (14 and 16).

(13) Jane was late for her appointment with Sue and she hurried to get a taxi.
(14) Jane was late for her appointment with Sue and hurried to get a taxi.
(15) Anna lent Felicity the steam iron and she forgot to give the instructions.
(16) Anna lent Felicity the steam iron and forgot to give the instructions.

Stevenson and Vitkovich found that when the verbs enabled pronoun resolution, subjects indicated the appropriate antecedent immediately; but when the verbs were uninformative, this decision was delayed until further downstream.

Finally, Garrod et al (1994) replicated the Garrod and Sanford (1985) study described in Chapter 3, but used measures of eye movements to indicate the immediate use of pronoun gender, discourse focus, and verb pragmatics to resolve pronominal reference. Subjects read a number of passages where potential antecedents could be distinguished by gender in one condition (17), but not in another (20). These passages were continued by a sentence making pronominal reference where the following verb matched the discourse role of one of the potential antecedents. In the following examples, *ordered* matches the discourse role of the passenger, Joan, while *poured* better fits the role of the steward(ess). Sentences (18) or (19) continued gender differentiated passages like (17), while one of (21) to (26) continued passages like (20); and both passages were completed by a final sentence (25).

(17) Flying to America
Joan$_1$ wasn't enjoying the flight at all. The dry air in the plane made her really thirsty. Just as she was about to call her, she noticed the stewardess$_2$ coming down the aisle with the drinks trolley.
(18) Right away she ordered$_1$ a large glass of coke.
(19) Right away she poured$_2$ a large glass of coke.

(20) Flying to America
Joan$_1$ wasn't enjoying the flight at all. The dry air in the plane made her really thirsty. Just as she was about to call him, she noticed the steward$_2$ coming down the aisle with the drinks trolley.
(21) Right away she ordered$_1$ a large glass of coke.
(22) Right away she poured$_2$ a large glass of coke.

(23) Right away he poured$_2$ a large glass of coke.
(24) Right away he ordered$_1$ a large glass of coke.

(25) Joan finished it in one go and ordered another one.

Garrod et al identified three referential constraints. The first of these is the discourse focus of the two potential antecedents, where the form and order of mention of characters determines which is focal (eg Sanford and Garrod, 1981), with Joan the focal character in (17) and (20) because she is mentioned first and by proper name. A second factor is pronoun gender, and a third is the pragmatic fit of the verb-description. Significant effects of pronoun and focus, and an interaction of these factors with verb pragmatics were observed on the verb (*ordered* or *poured*) using a first pass reading time measure[22], but not on the preceding pronoun. At the same time, all three factors were also seen to influence the total reading time of the anaphoric sentence. Garrod et al concluded that referential processes were triggered at the verb, and resolved further downstream, and that all three factors made an immediate impact on the referential process.

When Garrod et al replicated the experiment, but substituted names and definite descriptions for the pronouns, they found only total reading time indications of referential processes. Garrod et al concluded that anaphoric reference is immediately triggered on encountering pronouns, but not necessarily on encountering definite NPs or names. They suggest that this may be because a definite expression can either be interpreted anaphorically, or as the instruction to introduce a new element into the discourse representation. Garrod et al argue that these two possibilities are explored in parallel, but only selected between downstream of the expression.

This may explain the contradictory word recognition data, where definite NP reference appears to be more immediate than pronominal reference (eg Dell et al, 1983). Garrod et al suggest that the postulated anaphoric interpretation of a definite NP may access potential antecedents to the extent that their recognition

---

[22] Garrod et al used the same software to analyse their data as used in the present studies, and defined their measure of first pass reading time as the initial set of fixations on a given region of test, prior to rightwards eye movements to read subsequent regions of text, or leftwards eye movements to read previous regions of text. The first pass has generally been used as an indicator of immediate processing, and a description of early processes (Rayner and Pollatsek, 1989). However, its precise interpretation will depend on the size of the defined region of analysis. In the case of the Garrod et al materials these regions are small enough to warrant an immediacy interpretation. The total reading time measure records all the fixations on a text region, including those which are return visits to the region.

is facilitated in a recognition paradigm, but that fuller referential processing is delayed until the processor has settled on an anaphoric interpretation of the expression.

It may also account for Greene et al's failure to observe facilitated recognition of a pronoun's antecedent in a speeded reading and recognition task. If anaphoric processing is immediately initiated on encountering a pronoun, and this depends on a number of sources of information, then this will be both demanding on processing resources, and time-consuming. Since there is limited time to read the sentence in the Greene et al task, it seems likely that the on-line processing of anaphoric reference was disrupted, and subjects were unable to recover the pronoun's antecedent as they read the sentence. In contrast, if definite NPs do not trigger referential processes, but simply activate potential antecedents, then the corresponding anaphoric processing is not so easily disrupted by the speeded reading and recognition task.

Cloitre and Bever (1988) report some further experimental evidence that pronouns and definite NPs are used to access antecedents in different levels of the text representation. They found that when subjects read (or heard) short passages which ended with either an object pronoun or definite NP, there was a different pattern of either lexical or category judgements about a visually presented adjective that was earlier used to modify the appropriate antecedent. Subjects made faster lexical decisions following definite NP reference, but faster category membership decisions following pronominal reference. Cloitre and Bever concluded that pronouns directly access a conceptual or discourse representation of the antecedent, while definite NPs first access a surface lexical representation.

The present study followed Garrod et al in manipulating discourse focus and the pragmatic match between an anaphor and its antecedent. These two manipulations are highlighted in (26). According to my original predictions, monotone-increasing quantification will focus on the refset partition of the quantified NP, while monotone-decreasing quantification will result in focus on the compset partition. However, the SPR studies suggest that the refset and compset are equally accessible to pronominal reference following monotone-decreasing quantification. This may be because both of the subsets are in discourse focus, or because it is more easy to recover from anomalous reference following monotone-decreasing quantification. The present studies will determine whether these findings can be replicated.

(26) A Public Meeting.

Local MPs were invited to take part in a public enquiry about proposals to build a nuclear power station. *Quantifier* of the MP's attended the meeting. Their [*presence* / *absence*] helped the meeting run more smoothly.

The predictions are the same as those tested in the previous SPR studies, but with more precise claims about the locus of referential anomaly detection. There should be evidence of increased processing difficulty when reference is made to the unfocused subset of the quantified NP. This will be evident as an increased reading time on measures of First pass and Total reading times, which predicts a longer reading time for reference to the compset following quantification by *a few* or *many*, and for refset reference following quantification by either *few* or *not many*. The immediate detection of a referential anomaly should be observed an increased First pass reading time for the anaphoric NP. Failing this, there may be evidence of an increased number of regressions from this region of text.

Eye movements during reading

As already demonstrated by the Garrod et al study, an account of anaphoric reference which emphasises the initiation and completion of processing, and the integration of semantic and pragmatic constraints on resolution is amenable to the study of eye movements. It is also consonant with broader theoretical positions on the relationship between eye movements and cognitive processing.

Rayner and Pollatsek (1989) identify several theoretical stances. One position, which they describe as the **global control hypothesis**, proposes that eye movements proceed independently of cognitive processing. Other theories allow for the cognitive control of eye movements. Of these, the **immediacy hypothesis** proposed by Just and Carpenter (1980) makes the strongest claims. They argue that fixation is maintained on the currently processed word until processing of that word is completed. This means not only that visual and lexical processing is completed, but also syntactic and semantic processing, including the identification of referents. At the same time, Just and Carpenter allow for subsequent processing to revise initial and erroneous decisions. A less extreme position (Rayner and Pollatsek, 1989) argues that while eye movements are under linguistic control, they may not be rigidly shackled to current processing as supposed by Just and Carpenter. According to Rayner and Pollatsek, eye movements are likely to be under the direction of recently extracted text information, but not necessarily controlled by information extracted on the current fixation.

There is a body of evidence in support of the view that both eye movements and the duration of individual fixations are at least partly under cognitive control (see Rayner and Pollatsek, 1989, for a review). In terms of the evidence of eye movement control, a number of researchers (e.g. Rayner and McConkie, 1976; Rayner, 1979) have found that eye movements are sensitive to the length of individual words, and tend to be made within the beginning or middle regions of a word. Other studies have shown that, although short words (e.g. 3-letter words) tend not to be directly fixated, but are often 'skipped' during reading, those short words which are actually skipped also tend to have a high frequency of occurrence in language use, or are predictable from context.

It has also been widely demonstrated that the duration of individual fixations are under cognitive control. For instance, it has been widely reported (e.g. Inhoff and Rayner, 1986; Rayner and Duffy, 1986) that words which have a high frequency of occurrence in language are fixated with shorter latencies that those which have a low frequency of occurrence, even when controlling for word length. Longer fixation durations have also been reported for words which are syntactically ambiguous in the current sentence (Frazier and Rayner, 1982), and those which are lexically ambiguous (Rayner and Duffy, 1986). At the same time, shorter fixations are often made on words which are highly predictable in the current context (e.g. Ehrlich and Rayner, 1981).

The results for pronoun resolution and antecedent search are more pertinent to the present studies. Ehrlich and Rayner (1983) found that the total reading time of an anaphor was longer if its antecedent was more distant in the text. This measure of reading time included both the fixation time on first encountering the word and later re-fixations. In contrast, Garrod et al (1994) found evidence of an immediate increase in reading time on an anaphor and following verb when its antecedent is unfocused or pragmatically anomalous.

Further evidence suggests that while eye movements are under cognitive control, the immediacy hypothesis is too strong a claim. In particular, there is considerable evidence (e.g. Rayner, Well, Pollatsek and Bertera, 1982; Inhoff and Rayner, 1986) that more than the current word is processed on each fixation, and that the duration of individual fixations may not index the full processing of a word. Rayner et al (1982) found that subjects can partially process information from the word to the right of the currently fixated word during that fixation. Subjects read sentence under three conditions: one where only the currently fixated word was visible, but all the letters to the left of fixation were replaced by other letters; one where only the current word and the word to the left were visible, but all other letters were replaced by others; and one where the current word and part of the following word were visible. Rayner et al found that there was no difference in reading rate in conditions where the current and following word were visible, and others where the current word and part of the following word were visible, even when those visible letters in the following word were replaced by visually similar ones. In both cases the reading rate was faster than when there was no visible information about the following word. Rayner et al concluded that more than the currently fixated word can be processed on each fixation. The reader will also process information about the location and size of the following word. This information is used to program the next eye movement.

Inhoff and Rayner (1986) investigated the fixation duration on high and low frequency words in conditions where the word could or could not be pre-viewed prior to fixating on it. Inhoff and Rayner only observed a frequency effect on the first fixation when subjects were able to preview the word. However, there was evidence of a frequency effect when they totalled the durations of all the fixations on that word prior to an eye movement to fixate either the following or previous word. This suggested that when subjects were able to preview the following word, the following fixation was influenced by cognitive factors, but when there was no preview the first fixation was completed automatically and without attending to cognitive factors. However, if this first fixation was insufficient to complete processing, then the word was re-fixated.

The results from these two experiments are consistent with Rayner and Pollatsek's weaker version of the cognitive control hypothesis. There is substantial evidence that eye movements are guided by the processing of recently extracted text, but that there is not a rigid relationship between eye movements and the currently processed word. In particular, Inhoff and Rayner demonstrate that much of the processing of the current word is anticipated during the processing of the fixated word.

This complicates any attempt to draw inferences about cognitive processes from measures of eye movements. If there was a stronger relationship between eye movements and cognitive processes, then it would be possible to use the duration of the first fixation on a word as a index of processing. However, this is problematic since this word may have been processed to some extent in the previous fixation (e.g. Rayner et al, 1982), and that re-fixations on this word may contribute to even low-level word recognition (Inhoff and Rayner, 1986). An alternative is to use a measure of First pass reading time for a single word, which is the summation of all fixations made on that word prior to either a forwards eye movement in the text, or a regressive eye movement to re-read part of the previous text. However, this method can also be criticised. For instance, it is not clear that all of the re-fixations on the current word are made with the intention of continuing the processing of this word. They may be made in preparation for a saccade to fixate the next word. Neither is it possible to conclude that zero fixations on a particular word means that the word was not processed. Short function words, like the definite and indefinite articles, may be processed during a fixation on the prior word, and words may be 'skipped' because they are highly predictable from context and so processing is redundant.

This has proved problematic, and some researchers (e.g. Just and Carpenter, 1980) have been criticised for including the zero values in their analyses. However, this is not such a problem in the present studies, since the critical regions are generally quite long (2 words). Those trials where zero fixations are recorded on neighbouring regions of text are removed from the analysis since this means that the eye-tracker has failed to record fixations across a considerable length of text, and if this is not due to equipment error, it is reasonable to conclude that subjects are not reading the text as directed, i.e. they are not reading for comprehension.

Two further measures are reported in the following experiment. A measure of Total reading time, or a summation of all fixations spent within a region, including return visits to that region, is reported as a measure of global difficulty in processing the sentence. This may prove descriptive of processing which is not completed during the particular word. For instance, Ehrlich and Rayner (1983) report an differences in Total reading time as evidence of greater processing difficulty in anaphor resolution when the antecedent is distant in the text. Finally, a measure of regressive eye movements is reported, since while many of these will constitute corrections to word landing position, they may also be an index of processing difficulty and the re-fixation of previous text to overcome these difficulties (Rayner and Pollatsek, 1989).

Experiment three:

The time course of anaphoric reference to an NP quantified by *a few* or *few*.

Introduction

This first eye-tracking study was a replication of Experiment 1, which measured the reading time of plural anaphoric reference to an NP quantified by *a few* or *few*. The present study will provide a more fine-grained description of the time course of these processes.

Method

Design

The experiment used the same cross-over design as the previous self-paced reading studies. There were two with-subjects manipulations: the subject NP of the quantified sentence was quantified by either the monotone-increasing *a few* or the monotone decreasing *few*; and the following sentence began with an anaphor which described a property of either the refset or compset of the quantified sentence.

Subjects

Twenty-four subjects were used. All subjects were students at Glasgow University and were paid £5 for their participation. Prior to viewing the experiments materials, subjects were set-up on the eye-tracker and calibrated within a set criterion for recording accuracy. Subjects who either could not be set-up, or failed to make the criterion level during calibration were replaced and paid a lesser amount.

Other subjects were replaced following data collection, but prior to analysis, if there was missing data for more than 6 experimental items (25%). Failed items included those where no data was available because of either subject or eye-tracker error, and items where zero fixation time was recorded for either of two large text regions (one which contained a context-setting first sentence, and another which contained the quantified sentence), or two successive regions of text[23]. Again this was attributed to either subject or tracker error. Two subjects were replaced for these reasons.

Apparatus

A Stanford Research Institute Dual-Purkinje Eye-tracker was used to monitor eye movements. The eye-tracker has an angular resolution of 1′ arc. Although viewing was binocular, data was recorded from the right eye only.

Materials were presented on a VDU controlled by a Vanilla 386 computer. The subject's eye position was sampled every millisecond and analysed using software developed at the University of Massachusetts, and adapted by Keith Edwards at Glasgow University. The software continuously monitors the output to establish the sequence of fixations and measure their onset and end to the nearest millisecond.

---

[23] These text regions are defined in the Data analysis section.

Materials

The experiment used the 24 materials validated in the pilot study and used in the self-paced reading experiments reported in Chapter 4. Because the more fine-grained analysis afforded by the eye-tracking methodology also requires more stringent control of the materials, critical regions of the experimental materials were controlled for length and line position. The word length difference between the quantifiers could not be adjusted, but the other manipulated factor, the noun which qualified the plural pronoun was balanced as closely as possible across the two anaphoric conditions as closely as possible. It was a mean 6.83 characters long in the refset condition, and a mean 8.42 characters in the compset condition. Although these are not perfectly balanced, this was not considered to be a serious flaw, since cross-over effects were required to reject the null hypothesis. An example passage is given in (27).

(27) A Public Meeting.

> Local MP's were invited to take part in a public enquiry about proposals to build a nuclear power station. *Quantifier* of the MP's attended the meeting. Their [*presence* / *absence*] helped the meeting run more smoothly.

The monitor used to present the materials restricted line length to a maximum 65 characters, and the materials were arranged to accommodate this constraint while keeping the critical anaphoric noun-phrase section of the text in an approximately central line position. The pronoun began a minimum 15 (and a mean 19.79) characters from the start of the line to avoid landing errors associated with the return sweep from the end of the preceding line (Rayner and Pollatsek, 1989). Passages were presented with a single line spacing.

Materials were placed into four experimental blocks following a Latin square design. Six subjects viewed each of the experimental blocks, and subjects saw each material in only one of the experimental conditions.

There were also 24 filler items which were materials in another experiment. These were mixed with the experimental materials, and the full set of materials and filler items were divided into four presentation blocks with rest intervals between blocks. A further two practise items were added to the beginning of each block to familiarise subjects with the reading task. Comprehension questions requiring a YES or NO button press response followed 25% of the presented passages to ensure careful reading.

Experimental materials and filler items are listed in Appendix 1.

Procedure

Subjects read a sheet of instructions describing the general experimental procedure. They were then seated in front of the eye-tracker and the seat was adjusted for height and comfort. A head-strap and chin-rest were used to reduce head movement. Subjects fingers were placed on a key-pad with three keys: one to control the presentation of materials, and the others to give YES and NO responses to comprehension questions.

The eye-tracker was calibrated to the subject at the beginning of the experiment and following each of the 3 rest periods. A calibration check was made automatically before the presentation of each item, and subjects were re-calibrated if this failed.

During the experimental phase subjects read materials at their own pace.

Data analysis

There were three phases of data analysis. In the first phase, eye-fixation y-co-ordinates (fixations were analysed as x and y letter co-ordinates) were corrected by hand[24], before an automatic procedure summed short fixations within a given range of characters. All fixations separated by one character are pooled when where one or both of these fixations are less than 80 milliseconds. Fixations of less than 40 milliseconds are deleted from the eye movement record if they are separated by three or fewer characters from another fixation. These pooling procedures worked under the assumption that adjacent fixations of short duration represent the same fixation separated by a micro-saccade (see Rayner and Pollatsek, 1989). Following the data-correction procedures, measures of eye-movement and fixations were calculated across subjects and materials for specified region divisions.

The materials were first of all divided into sentence regions to produce data which was comparable with that from the SPR studies. The anaphoric sentence was then divided into smaller regions in order to produce a more detailed

---

[24] Because the eye-movement output file is encoded as x and y letter co-ordinates, y fixations which land towards the top or bottom of letters was sometimes encoded with the value of a neighbouring line (ie fixations on text at line 8 could be encoded as line 7 or 9). These errors were corrected by hand, following the rule that fixations could 'wander' by up to two lines, and those fixations which monotonely increased along the x-axis should be interpreted as a series of fixations with the same y co-ordinate value.

account of the time course of processing. The regions used for this analysis are illustrated in (28). Region 1 is the context region, and contains all of the text prior to the quantifier manipulation. Region 2 is the quantified sentence. The final sentence was divided into three regions, where region 3 contains the anaphoric noun-phrase manipulation, region 4 is the subsequent verb, and region 5 contains the remainder of the sentence.

(28) **R1:** A Public Meeting.

Local MP's were invited to take part in a public enquiry about proposals to build a nuclear power station.

R2: *Quantifier* of the MP's attended the meeting.

R3: Their *presence* / *absence*

R4: helped

R5: the meeting run more smoothly.

Eye-movement measures for the quantified sentence allowed an analysis of the processing differences between monotone-increasing and monotone-decreasing quantification suggested by the previous experiment, and verification studies (see Chapter 1 for a review of verification studies). Region 3 is the critical measurement region for anaphoric reference, while region 4 was measured separately to record any limited spill-over effects (Ehrlich and Rayner, 1983; Rayner and Pollatsek, 1989; Garrod et al, 1994).

Reported eye-tracking measures

Three eye-movement measures are reported. Two of these are measures of reading time[25]: which is defined here as the time spent fixating on a piece of text not including the time required for eye movements. The **Total reading time** is a measure of reading time for all visits to a region of text, including return visits from downstream regions. An increase in reading time for this measure indicates either that extra processing is required to overcome a problem encountered at the current region, or that information is sampled from the region to solve a processing problem encountered at some point beyond the current region. These can be distinguished using results from the other measures. The Total reading time for a full sentence is comparable with the reading time data from the SPR studies, although it should be remembered that subjects cannot return to previously read sentences in the SPR experiment reported in this thesis, but can do in an eye movement study.

[25] Measured here as milliseconds per character, a correction for differences in region size (but see Tannenhaus, Trueswell and Garnsey, 1994).

First pass reading time is a measure of the time spent fixating in a region of text before movement in either direction out of that region, and is generally regarded as the best measure of immediate processing (Rayner and Pollatsek, 1989). Rayner and his colleagues (eg. Rayner and Frazier, 1987) use the First pass measure to distinguish between initial syntactic processing and higher-level processing. As we are implicitly questioning this distinction, it seems reasonable to use the same measure to locate semantic and pragmatic influences on processing in the first pass. There are precedents for this approach. Garrod et al (1994) found evidence of immediate discourse focus and pragmatic influences on referential processing using this measure; while Garrod et al (1990) used it to detect on-line instrumental inference.

The third measure reported is the **First pass regressions** for a region. These are defined as any right to left eye movements made from the leading edge fixation in a region, i.e. the most rightward fixation that has so far been made in the region. The analysis included regressions made to other points within, and to points outside of the current region. Although the within-region regressions will include eye movements to accurately target a word (Rayner and Pollatsek, 1989), others may represent back-tracking in response to current processing difficulties. It is assumed the corrective fixations will occur randomly across experimental conditions.

Results

Trials were removed prior to analysis of the First Pass or Total reading time if zero scores were recorded for the context or quantified sentence regions, or for neighbouring regions. Zero scores were either due to tracker loss, or subjects not fixating within the region. Although it could be claimed that subjects had made a deliberate decision not to spend any time in a particular region (and so return a zero value), since the present regions were fairly long, a zero score for two regions meant that the subject had skipped at least three words. This degree of word-skipping would suggest that subjects were not reading, or comprehending, the text as required. Approximately 5% of trials were removed prior to an analysis of the Total reading time, while 7% were lost prior to the First pass reading time. This difference arises from a greater tendency to skip regions in the First pass analysis. These regions are often returned to in a second pass.

Sentence reading time

The Total reading time data was first of all analysed as a sentence reading time for the context, quantified and anaphoric sentences. This produced data which is comparable with SPR reading times. The data for the three sentences was separately analysed using 2x2 within subject ANOVAs.

The anaphoric sentence analysis produced a significant interaction of quantifier and reference types ($F1(1,23)=33.86$, $p<0.001$; $F2(1,23)=15.56$, $p<0.01$), with longer reading times for reference to the compset partition of an NP quantified by *a few*, and reference to the refset of one quantified by *few*. The mean Total reading time for the four experimental conditions is illustrated in Figure 1.

Further analyses of simple effects confirmed that both quantifiers were responsible for the interaction. There was a significant difference in reading time between refset and compset reference following quantification by *a few* ($F1(1,23)=21.52$, $p<0.001$; $F2(1,23)=9.80$, $p<0.005$); and between refset and compset reference following quantification by *few* ($F1(1,23)=12.90$, $p<0.002$; $F2(1,23)=5.99$, $p<0.023$). At the same time, the two anomalous reference conditions (compset reference following quantification by *a few*, and refset reference following quantification by *few*) were of the same magnitude, as were the contrasting felicitous reference conditions (all $Fs<1.0$).

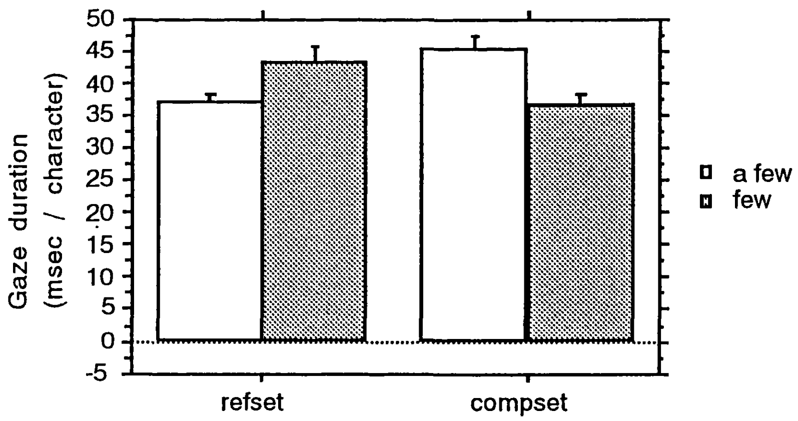


Figure 1: mean total reading time (with standard error bars) for the anaphoric sentence following quantification by *a few* and *few*

No significant effects were found for analyses of the context or quantified sentences ($Fs<2.2$).

Total reading time for regions of the anaphoric sentence

Separate 2x2 within-subjects ANOVAs were used to analyse each of the region division of the anaphoric sentence, and revealed significant interactions of monotonicity and reference type on the anaphor, verb and final regions of analysis. The mean reading times for all five regions of analysis (including the context and quantified sentences) are illustrated in figure 2.

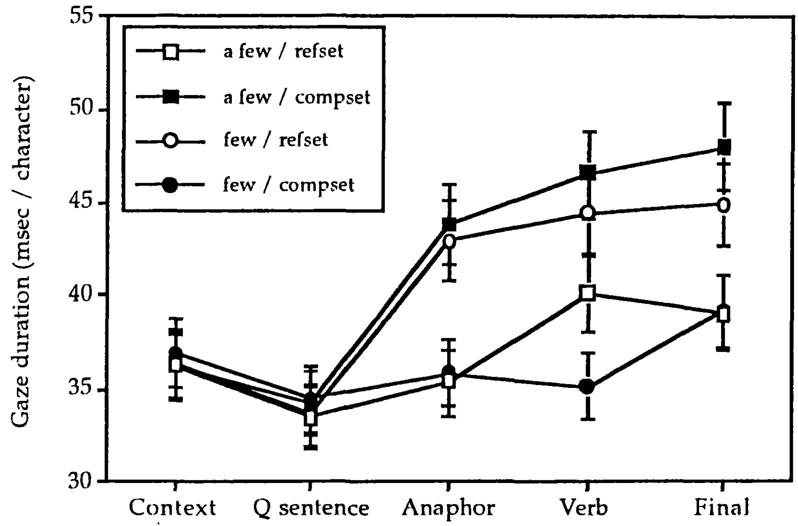


Figure 2: Mean total reading time (with standard error bars) following quantification by *a few* and *few*

Anaphoric region

The interaction of quantifier and reference type ($F1(1,23)=14.71$, $p<0.001$; $F2(1,23)=7.33$, $p<0.013$) followed the same pattern at the sentence reading time results, with longer reading times for the two conditions making anomalous reference to the quantified NP.

An analysis of the simple effects contrasts for the anaphoric region confirmed that the interaction was due to significant differences between refset and compset reference in the two quantificational conditions. The difference in reading time between compset and refset reference following quantification by *a few*, was significant by subjects, and marginal by items ($F1(1,23)=7.75$, $p<0.011$; $F2(1,23)=3.60$, $p<0.071$); as was the difference in reading time between these reference conditions following quantification by *few* ($F1(1,23)=6.98$, $p<0.015$; $F2(1,23)=3.73$, $p<0.067$). There was no reliable difference between the two felicitous referential conditions ($Fs<0.1$), or between the two anomalous reference conditions ($Fs<0.1$).

Verb region

The interaction of quantifier and reference type ($F1(1,23)=13.79$, $p<0.002$; $F2(1,23)=7.98$, $p<0.010$) for this region was again due to longer reading times for the two anomalous reference conditions. An analysis of the simple effects confirmed that this was due to both quantifiers. There was a significant difference in reading time between refset and compset reference following quantification by *a few* which was significant by subjects and marginal by items ($F1(1,23)=6.53$, $p<0.018$; $F2(1,23)=4.20$, $p<0.053$), and a significant difference following quantification by *few* which was also significant by subjects and marginal by items ($F1(1,23)=7.27$, $p<0.013$, $F2(1,23)=3.79$, $p<0.065$). This suggests that both quantificational conditions contributed to the interaction. There was no difference between the two felicitous reference conditions ($Fs<1.0$) or between the two anomalous reference conditions ($Fs<1.0$), which suggests that felicitous reference is equally easy, and anomalous reference equally difficult, following the two types of quantification.

Final region

The interaction of quantifier and reference type ($F1(1,23)=25.43$, $p<0.001$; $F2(1,23)=10.32$, $p<0.004$) for this region had the same pattern again: longer reading times for anomalous reference to the quantified NP. An analysis of simple effects found significant differences in reading time for refset and compset reference following quantification by *a few* ($F1(1,23)=16.99$, $p<0.001$; $F2(1,23)=7.43$, $p<0.013$); and a significant difference between these reference types following quantification by *few* on a subjects analysis only ($F1(1,23)=9.07$, $p<0.007$; $F2(1,23)=3.30$, $p<0.083$). There was no difference between the two felicitous reference conditions ($Fs<1.0$), and a difference between the two anomalous reference conditions which failed to reach significance ($F1(1,23)=2.42$, $p<0.134$; $F2(1,23)=1.11$, $p<0.303$).

Summary of total reading time results

The Total reading time measures were consistent with the original predictions. There was a longer reading time for reference to the unfocused subset of a quantified NP than for reference to the focal subset; which meant that there were longer reading times for reference to the compset following quantification by *a few*; and for refset reference following quantification by *few*. This pattern was evident in both an analysis of the Total reading time for the anaphoric sentence, and for more fine-grained analyses of component regions. Furthermore, both sets of data established that there was no difference in reading time between the two felicitous reference conditions, or between the two anomalous reference conditions, which suggests that the processing of anomalous reference is equally difficult, and the processing of felicitous reference equally easy, following monotone-increasing and monotone-decreasing quantification.

First pass reading time

The First pass reading time data for the five regions was separately analysed using a 2x2 within subjects ANOVA. The mean First pass reading times for all five regions is illustrated in Figure 3.

There was an interaction of quantifier and reference type in the predicted direction at the final region which was marginal by subjects ($F1(1,23)=4.19$, $p<0.052$; $F2(1,23)=2.39$, $p<0.136$). An inspection of Figure 3 suggests that this effect is principally due to a longer reading time following reference to the compset of an NP quantified by *a few*. There was also a main effect of reference type for the quantified sentence ($F1(1,23)=8.07$, $p<0.001$; $F2(1,23)=5.55$, $p<0.028$), with a longer reading time for those conditions making reference to the compset - although it must be noted that subjects had not yet encountered the region of text which made this reference. No other effects were observed on any of the regions (all $Fs<1.3$).

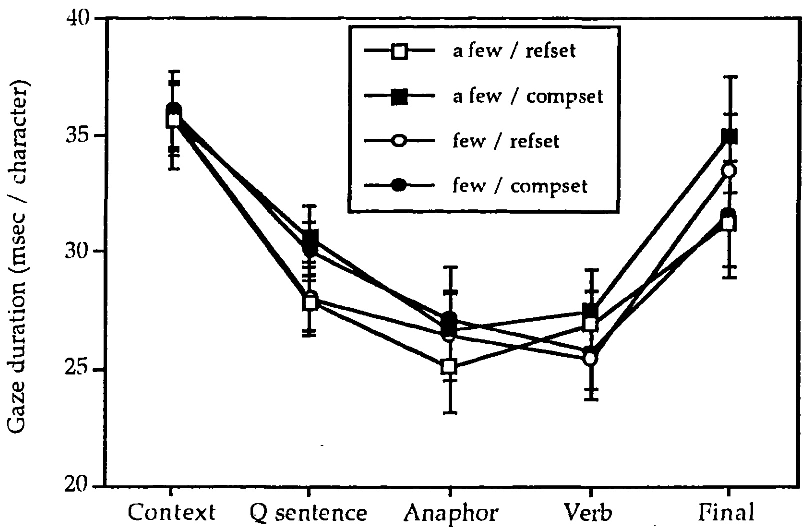


Figure 3: mean first pass reading time (with standard error bars) following quantification by *a few* and *few*

First pass regressions

As before, the regressions made from each region were separately analysed using 2x2 within subjects ANOVA's. However, these analyses only considered the anaphor, verb and final regions, as first pass regressions from the quantified sentence or context region are not informative about the experimental manipulation. Effects were observed on the anaphor and verb regions only (all other $Fs<1.9$); and are illustrated in Figures 4 and 5 respectively.

Anaphor region

An analysis of the anaphor region produced a main effect of quantifier type which was significant by subjects, and marginal by items ($F1(1,23)=7.04$, $p<0.014$; $F2(1, 23)=3.99$, $p<0.058$). More regressions were made in the monotone-decreasing *few* conditions.

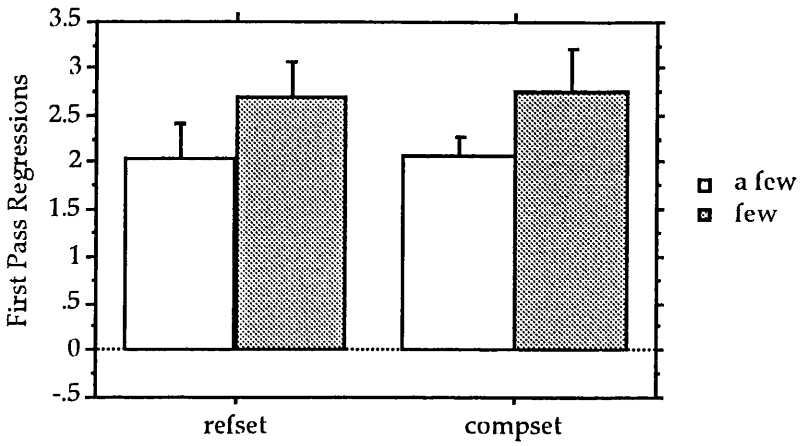


Figure 4: First pass regressions from the anaphor region (with standard error bars) following quantification by *a few* and *few*

Verb region

The verb region produced a marginal main effect of quantifier type by subjects, which was significant on an items analysis, and marginal by subjects ($F1(1,23)=3.50$, $p<0.074$; $F2(1, 23)=4.15$, $p<0.033$); and an interaction of quantifier and reference type ($F1(1,23)=5.99$, $p<0.0224$; $F2(1, 23)=6.14$, $p<0.020$). An inspection of the means (see Figure 5) suggests that this was due to a greater number of regressions for compset reference following quantification by *a few*. An analysis of simple effects confirmed this contrast. There was a significant difference between refset and compset reference following quantification by *a few* ($F1(1,23)=7.25$, $p<0.013$; $F2(1,23)=8.09$, $p<0.009$); a significant difference between the two anomalous reference conditions ($F1(1,23)=5.93$, $p<0.023$; $F2(1,23)=6.80$, $p<0.0160$); and between the two felicitous reference conditions ($F1(1,23)=10.27$, $p<0.004$; $F2(1,23)=11.01$, $p<0.003$).

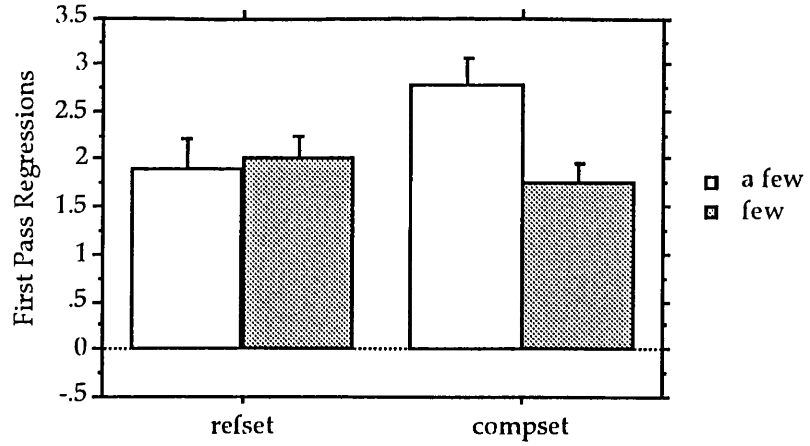


Figure 5: First pass regressions from the verb region
(with standard error bars) following quantification by *a few* and *few*

Summary of first pass results

There was no clear evidence of immediate anomaly detection according to measures of either First pass reading time or First pass regressions. However, there was evidence of an overall increase in regressions from the anaphor region following quantification by *few*, and an increased number of regressions from the following verb region when reference was made to the compset following quantification by *a few*. Since an increase in regressions can be interpreted as evidence of processing difficulty, it can be concluded that both of these findings indicate some difficulties in processing reference during the first pass of the sentence. In particular, there appears to be a general difficulty in processing reference following monotone-decreasing quantification by *few*; and some evidence that reference to the compset is more difficult following monotone-increasing quantification by *a few*. There was also some evidence that this difficulty in processing reference to the compset is reflected in the First pass reading time for the final region of the anaphoric sentence.

Discussion

The results from this first eye movement experiment were consistent with the findings of the previous self-paced reading experiments. That is, there was evidence of an increased reading time, and therefore processing difficulty, for anaphoric sentences which referred to a property of the unfocused subset.

According to both measures of Total reading time for the full sentence, and for regions of that sentence, there was an increased reading time when reference was made to the compset of a sentence quantified by *a few*, and reference to the refset of a sentence quantified by *few*. The analysis of sentence regions suggested that this disruption was localised to the anaphor and verb regions following quantification by *few*.

However, there was no convincing evidence that the anomalous reference was detected on-line, i.e. concurrently with the processing of the anaphor. There was some evidence from measures of First pass regressions that anaphoric reference was more difficult, and therefore resulted in more regressions, following quantification by *few*. Furthermore, there was evidence of increased regressions from the verb region, and a suggestion of an increased First pass reading time for the final region following reference to the compset of a sentence quantified by *a few*.

While these results can be interpreted as evidence of a disruption of processing during the first pass of a sentence when reference is made to an unfocused subset, these results were not predicted in advance, so must be treated with caution. Moreover, the first pass reading time results were complicated by a highly significant main effect of reference type at the quantified sentence, which was prior to the reference manipulation. This must be a spurious effect, because it is not based on any information available at that point in time. There was no evidence of a difference in the processing of the two forms of quantified sentence according to the measure of Total reading time.

In summary, the strongest conclusions that can be drawn is that there is a sentence-level difficulty in integrating the anaphoric sentence with prior discourse when the anaphor describes a property of the unfocused subset of the preceding quantified sentence. However, further conclusions about the on-line processing of reference are possible if the present pattern of First pass results are replicated. For that reason, a fourth experiment examines the processing of reference to a sentence quantified by either *many* or *not many*.

Experiment four:

The time course of anaphoric reference to an NP quantified by *many* or *not many*.

Introduction

This experiment was a replication of the previous study, with the substitution of the quantifiers *many* and *not many* for *a few* and *few* respectively.

Method

The experiment following the same design and procedure as Experiment three, and used the same set of materials. The only difference was that the passages were double-spaced in the presentation, which made easier the initial stages of data analysis.

Subjects

Twenty-four subjects were used for this study and paid £5 for their participation. Subjects who failed to calibrate were replaced and paid a lesser amount. Another four subjects were replaced prior to an analysis of the results because there was more than 25% data loss due to subject or tracker errors.

Materials

The experimental materials were divided into the same text regions as used in the previous experiment. These are illustrated in (29). The first region contained all of the text until the quantified sentence, and the second region contained the quantified sentence. The anaphoric sentence was divided into three regions. Region 3 contained the anaphor, region 4 contained the following verb, and region 5 contained the remainder of the sentence.

(29) **R1:** A Public Meeting.

Local MP's were invited to take part in a public enquiry about proposals to build a nuclear power station.

R2: *Quantifier* of the MPs attended the meeting.

R3: Their *presence / absence*

R4: helped

R5: the meeting run more smoothly.

Results

As with the previous experiment, trials were removed prior to analyses of Total Time and First pass reading times if no fixations were made on the context region, quantified sentence, or two successive regions. These are due to tracker loss, or subjects' failing to fixate in the region. Approximately 5% of data were removed prior to the Total reading time analysis, and approximately 6% of data were removed prior to the First pass reading time analysis. As before, the difference between these amounts reflects a greater tendency to skip words during a first pass of the text.

Sentence reading time

The Total reading time data was first of all analysed as a sentence reading time for the context, quantified and anaphoric sentences using separate 2x2 within subject ANOVAs.

An analysis of the anaphoric sentence produced a significant main effect of reference type by subjects only ($F1(1,23)=6.73$, $p<0.017$; $F2(1,23)=2.29$, $p<0.144$), which was due to longer reading times for those conditions which described the compset partition of the quantified NP; and a significant interaction of quantifier and reference type ($F1(1,23)=19.19$, $p<0.001$; $F2(1,23)=16.39$, $p<0.001$). This was due to longer reading times for conditions making anomalous reference to the quantified NP. The mean Total reading times for the four experimental conditions is illustrated in Figure 6.

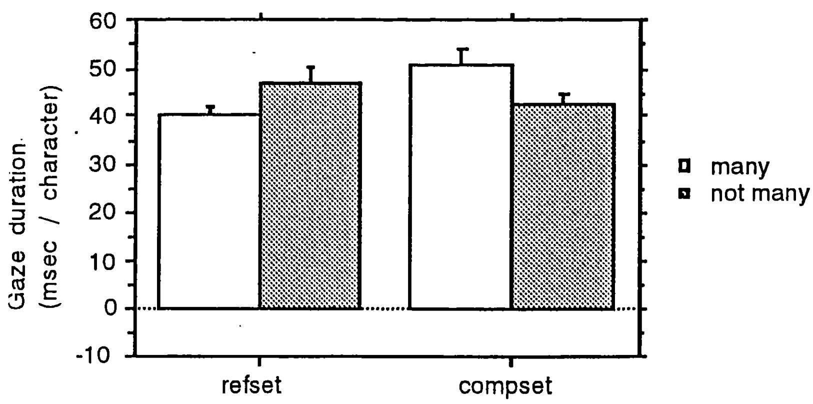


Figure 6: Mean total reading time (with standard error bars) for the anaphoric sentence following quantification by *many* and *not many*

An analysis of the simple effects suggested that quantification by *many* was primarily responsible for the interaction. There was a significant difference between refset and compset reference following quantification by many ($F1(1,23)=9.25$, $p<0.001$; $F2(1,23)=15.86$, $p<0.001$); but only a marginal difference between the reference conditions following quantification by *not many* (F1(1,23)=3.27, p<0.084; F2(1,23)=3.04, p<0.095). However, there was no difference between the two anomalous reference conditions, or between the two felicitous reference ones (all Fs<1.0).

The analyses of the other sentence reading times produced effects which were significant on the subjects analysis only. An analysis of the context region produced an interaction of quantifier and reference type in the predicted direction, which was marginal on the subjects analysis and non-significant by items (F1(1,23)=3.43, p<0.077; F2(1,23)=0.33, p<0.575). Similarly, an analysis of the quantified sentence produced an interaction in the predicted direction which was only significant on the subjects analysis (F1,23)=4.01, p<0.057; F2(1,23)=1.492, p<0.234). There were no other significant effects (Fs<1.0).

Total reading time for regions of the anaphoric sentence

Separate analyses of the total reading time were carried out for sub-regions of the anaphoric sentence using separate 2x2 within-subjects ANOVAs. Figure 7 illustrates the mean Total reading times for these regions, and for the contextual and quantified sentences.

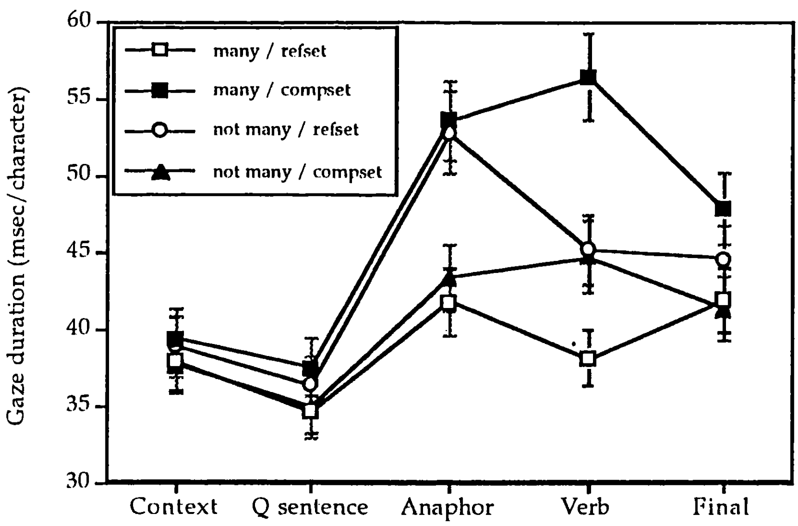


Figure 7: Mean total reading time (with standard error bars) following quantification by *mankind not many*

Anaphor region

There was a significant interaction of quantifier and reference type in the predicted direction for the anaphor region (F1(1,23)=14.35, p<0.001;

F2(1,23)=10.16, p<0.005). This is illustrated in Figure 10 as a marked increase in reading time for the two anomalous reference conditions.

An analysis of simple effects means contrast confirmed that both quantifier manipulations were driving this interaction. There was a significant difference between refset and compset reference conditions following quantification by *many* (F1(1,23)=8.83, p<0.007; F2(1,23)=5.78, p<0.025); and between the two reference conditions following quantification by *not many* (F1(1,23)=5.69, p<0.026; F2(1,23)=4.43, p<0.047). Furthermore, there were no significant differences between the two felicitous reference conditions (Fs<1.0), or between the two anomalous reference conditions (Fs<1.0).

Verb region

The analysis of the verb region also produced a significant interaction of quantifier and reference type (F1(1,23)=11.38, p<0.003; F2(1,23)=21.34, p<0.001). An inspection of Figure 10 suggests that a difference between the two reference conditions following quantification by *many* is responsible for this effect, and that there is no difference between the two reference conditions following quantification by *not many*. This was confirmed by subsequent simple effects means comparison tests.

There was a significant difference between refset and compset reference after quantification by *many* (F1(1,23)=21.54, p<0.001; F2(1,23)=36.62, p<0.001); but no significant difference between the reference conditions following quantification by *not many* (F1(1,23)=0.017, p<0.898; F2(1,23)=0.23, p<0.635). At the same time, there was no significant difference between the felicitous reference conditions (F1(1,23)=0.68, p<0.149; F2(1,23)=2.22, p<0.150), but a significantly shorter reading time for anomalous reference following quantification by *not many* compared to anomalous reference following quantification by *many* (F1(1,23)=4.90, p<0.038; F2(1,23)=6.04, p<0.023).

Final region

Finally, an analysis of the final region produced results where an interaction in the predicted direction approached significance by both subjects and items (F1(1,23)=3.29, p<0.083, F2(1,23)=3.11, p<0.092).

There were no other significant effects on analyses of any regions (all Fs<1.0).

Summary of the Total reading time results

The analysis of Total reading time supports the claim that there are global difficulties in processing the anaphoric sentence when reference is made to the unfocused subset. According to the analysis of sentence reading time, there was a longer reading time, and therefore greater processing difficulty, when reference was made to the compset as compared to the refset following quantification by *many.* However, there was no significant difference between the two reference conditions following quantification by *not many.*

On the surface, this suggests that reference was equally easy (or difficult) following quantification by *not many.* However, an analysis of Total reading time across the regions of text demonstrated that there was a longer reading time at the anaphor region when it referred to a property of the refset. It appears that the increased difficulty is a localised effect which is barely detectable at the level of sentence-reading time.

The analysis of Total reading time across the sentence regions confirmed the greater difficulty in processing reference to the compset following quantification by *many,* with longer reading times for both the anaphor and verb regions compared to refset reference.

First pass reading time

There were no significant effects at any region on an analysis of first pass reading time (all $Fs<1.0$).

First pass regressions

As with experiment 2a, we analysed the first pass regressions for regions 2 to 5 using separate 2x2 within-subjects ANOVAs.

Only an analysis of the final region produced significant results. There was a main effect of reference type which was significant by subjects only ($F1(1,23)=8.71$, $p<0.008$; $F2(1,23)=2.41$, $p<0.135$) and an interaction of quantifier and reference type in the predicted direction ($F1(1,23)=8.13$, $p<0.009$; $F2(1,23)=9.91$, $p<0.005$). Inspection of the results (see Figure 8) shows that the effect is principally due to a greater number of regressions for compset reference following quantification by *many,* which was confirmed by an analysis of simple effects. There was a significant difference between the two reference conditions following quantification by *many* ($F1(1,23)=14.85$, $p<0.001$; $F2(1,23)=11.30$, $p<0.003$), between the two compset reference conditions ($F1(1,23)=8.23$, $p<0.009$; $F2(1,23)=8.10$, $p<0.01$), and between the *many /*

compset reference and *not many* / refset reference conditions on the subjects analysis only (F1(1,23)=7.23, p<0.014; F2(1,23)=3.07, p<0.094).

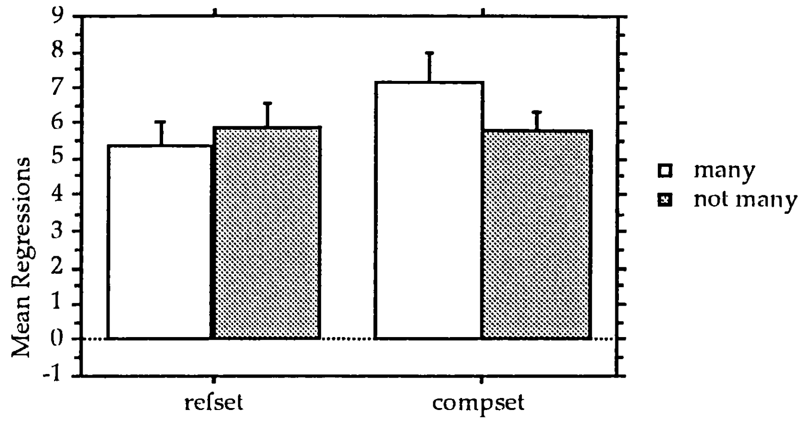


Figure 8: First pass regressions from the final region (with standard error bars) following quantification by *many* and *not many*

Summary of first pass results
There was no evidence of on-line anomaly detection according to either measure of First pass reading time or regressions. That is , there was no evidence on an increased reading time or increased number of regressions on the first pass through the anaphor region. However, the analysis of First pass regressions did provide some evidence of a disruption to processing during the first pass through the sentence when reference was made to the compset following quantification by *many*. There was a greater number of regressions from the final region of text in this compared to the other experimental conditions.

Discussion

As before, the results confirm that there are global difficulties in processing the anaphoric sentence when it made reference to the unfocused subset of a quantified sentence. However, while in the previous experiment this was evident from the Total reading time for the sentence, a more fine-grained analysis of the Total reading times for individual test regions proved necessary in the current experiment. A measure of sentence reading time suggested that, while there was a longer reading time in processing reference to the compset of a sentence quantified by *many*, there was no difference in reading time for reference to either the refset or compset following quantification by *not many*. However, an inspection of the Total reading times for individual text regions demonstrated that there was a longer reading time at the anaphor region for reference to the refset compared to the compset following quantification by *not many*. This finding is both a further demonstration that reference to an unfocused subset results in global processing difficulties, and that fine-grained measures of eye movements can detect short-lived and localised effects which cannot be observed on coarser measures of sentence reading time.

There was also some evidence that the context and quantified sentences was often re-read following anomalous reference, although these effects were only significant on a subjects analysis.

As with the previous eye movement experiment, there was no evidence of on-line referential anomaly detection on the measure of First pass reading, meaning that anomalous reference produced either an increased First pass reading time or number of regressions as subjects read the anaphor. However, there was no evidence of the spurious effect observed at the quantified sentence in that experiment either.

There was some evidence of a processing disruption during the first pass of the anaphoric sentence when reference was made to the compset of a sentence quantified by *many*. This was found in an analysis of regressions. There was a significantly greater number of regressions from the final region.

In summary, there was further evidence that the anaphoric sentence is more difficult to integrate with previous discourse when it refers to a property of the unfocused subset of the preceding quantified sentence. There was no evidence that this anomaly was detected on-line, although there was some evidence that processing was disrupted during the first pass of the sentence when reference was made to the unfocused compset following quantification by *many*.

A combined analysis of Experiments 3 and 4

The First pass reading time and regressions data were combined in order to increase the power of the analysis. This would increase the likelihood that any weak effects would be detected. If this is the case, then there should be evidence of two-way interactions of monotonicity and reference type for either reading time or regressions within individual text regions. In contrast, three-way interactions of experiment, monotonicity type and reference will be due to effects for particular pairs of quantifier (*a few* & *few* or *many* & *not many*) and may have been found in analyses of the separate experiments.

First pass reading time

The First pass reading time results for the four quantifiers across the five regions were analysed separately using a 2 (experiment) x 2 (quantifier monotonicity) x2 (reference) mixed design ANOVA. Only an analysis of the final region produced an effect. There was a three-way interaction of experiment, monotonicity and reference type, which was significant by subjects and marginal by materials ($F1(1,46)=5.15$, $p<0.029$; $F2(1,46)=3.83$, $p<0.057$). This effect is restricted to the *a few* & *few* quantifier pair, and due to a longer reading time for compset reference following quantification by *a few*.

First pass regressions

The first pass regressions were also analysed using a 2 (experiment) x 2 (quantifier monotonicity) x2 (reference) mixed design ANOVA. An interaction of experiment, monotonicity and reference type was observed on the anaphor region and was significant by subjects and marginal by materials ($F1(1,46)=4.73$, $p<0.035$; $F2(1,46)=3.800$, $p<0.058$). The same interaction was observed on the following verb region, but failed to reach significance ($F1(1,46)=2.93$, $p<0.092$; $F2(1,46)=1.72$, $p<0.197$), while an interaction of quantifier and reference type was significant by materials only ($F1(1,46)=2.56$, $p<0.117$; $F2(1,46)=4.44$, $p<0.041$). Finally, a main effect of reference type was observed on the final region ($F1(1,46)=8.67$, $p<0.006$; $F2(1,46)=4.17$, $p<0.047$), which is due to an overall increase in regressions following reference to the compset. There was also a significant interaction of monotonicity and reference type on the materials analysis only for this region ($F1(1,46)=2.68$, $p<0.109$; $F2(1,46)=11.170$, $p<0.002$). No other significant effects ($Fs<2.8$) were observed.

Discussion

The combined analysis failed to uncover any effects that were not already described by the analyses of the individual experiments. This suggested that there was no common patterns of processing during the First pass of the anaphoric sentence for either the two monotone-increasing or monotone-decreasing quantifiers, which invites the conclusion that quantifier monotonicity and contingent reference are not processed during the first pass.

Summary of the experimental findings

The eye-tracking experiments provided some further support to the claim that monotone-increasing and -decreasing quantifiers exhibit different patterns of focus license reference to different subset. There was evidence of an increased Total reading time for reference to a property of the compset of a monotone-increasing quantified sentence, and for reference to a property of the refset of a monotone-decreasing quantified sentence. Since increased Total reading time can be interpreted as evidence of difficulty in processing anaphoric reference (Ehrlich and Rayner, 1983), it was possible to conclude that reference was more difficult in these cases.

However, there was no evidence that anomalous reference was detected on-line, i.e. concurrently with the processing of the anaphor, although the analysis of First pass reading time and regressions suggested that there was some disruption of processing some distance downstream of the referential manipulation. However, a combined analysis of these reading time and regression results failed to identify any common pattern of results, which suggests that monotonicity alone cannot account for the pattern of disruption observed in the individual experiments. This suggests that the results were either spurious, or that the disruption depends on individual properties of the quantifiers. The most reasonable conclusion is that referential anomalies are not detected in the first pass. The following chapter reviews the results from the four experiments reported so far, and considers why there the referential anomaly is not detected on-line.

Chapter six:

A discussion of the experimental findings.

Introduction

This chapter will take some time out to review the experimental findings so far, and to consider what they add to our understanding of natural language quantification and referential processing. It will become clear that the studies invite more questions than they answer. In particular, it turns out that the use of eye-movement measures to identify the locus of referential anomaly detection only illustrates the complexity of this question, instead of providing an unequivocal answer. The remainder of the chapter considers the processing of referential anomaly detection in more depth, and suggests several modifications to the experimental design which might prove more informative about the time course of these processes. The next chapter reports some favourable results from an experiment which employed some of these modifications.

Results from the SPR and eye movement studies

Both the self-paced reading (SPR) experiments reported in Chapter 4, and the eye movement experiments reported in Chapter 5, support the hypothesis that some quantifiers are used to focus on and enable reference to different subsets of the quantified sentence. In particular, monotone-increasing quantifiers like *a few* and *many* focus on and enable reference to the refset partition, but block reference to the compset. In fact, given the difficulty experienced when reference is made to the compset, it seems likely that the compset is not mentally represented following monotone-increasing quantification. Monotone-decreasing quantifiers like *few* and *not many* appear to have a more diffuse pattern of focus, where both the refset and compset can be referred to, but the compset is the preferred antecedent.

The first SPR experiment measured the reading time for anaphoric sentences which referred to a property of the refset and compset partition of a sentence quantified by either *a few* or *few*. There was an increased reading time for reference to the compset as compared to the refset following quantification by *a few*; but no reliable difference between these two forms of reference following quantification by *few*. The second SPR experiment substituted *many* for *a few*, and *not many* for *few*. Again there was an increased reading time for reference to the compset as compared to the refset following quantification by *many*, but also an increased reading time for refset as compared to compset reference following quantification by *not many*.

Since measures of sentence reading time can be used as an index of the ease of integrating the current sentence with prior discourse (Haviland and Clark, 1974), and more specifically the ease of processing anaphoric reference (Garrod and Sanford, 1977), it was possible to conclude that the increased reading times in the above examples were indicative of difficulties in referential processing in different quantificational conditions. That is, subjects found it difficult to process reference to the compset of a monotone-increasing quantified sentence (i.e. one quantified by either *a few* or *many*). However, there was a less reliable pattern of results following monotone-decreasing quantification. It appeared that reference to the refset or compset were equally easy (or difficult) to process following quantification by *few*; but that refset reference was more difficult following quantification by *not many*.

The eye movement experiments produced a similar pattern of results on measures of Total reading time across either the length of the anaphoric sentence, or sub-regions of text. On a comparison of reference to sentences quantified by *a few* or *few*, there was a longer Total reading time on the anaphor and following verb region for reference to the compset as compared to the refset following quantification by *a few*. Conversely, there was a longer Total reading time on the anaphor and verb regions when reference was made to the refset following quantification by *few*. The fourth experiment compared reference to sentences quantified by *many* and *not many*. This time there was an increased reading time at the anaphor and verb regions for reference to the compset following quantification by *many*, and at the anaphor region for refset reference following quantification by *not many*.

Interestingly, there was no evidence of a difference in reading time for the entire anaphoric sentence when reference was made to either the refset or compset following quantification by *not many*. This suggested that reference to an unfocused subset appears less anomalous, and more easily recovered from, following monotone-decreasing quantification. This is probably because both the refset and compset are plausible antecedents under these conditions. It may also explain the failure to find a difference in self-paced sentence reading time following quantification by *few*. If subjects found the experience of anomalous reference to be short-lived and easily recovered from in this experiment, then the self-paced reading methodology may have been too course-grained to detect this difference.

The time course of processing

So far we have considered the relative magnitude of the reading time for anaphoric reference to monotone-increasing and monotone-decreasing quantified NPs. However, much of the reason for conducting eye movement studies was to gain a more fine-grained understanding of the time course of this processing, and the locus of referential anomaly detection. It should be possible to tell from this whether anaphoric reference to a quantified NP is computed on-line, and if there are any differences in the time course of processing for monotone-increasing and monotone-decreasing quantifiers.

I argued in the introduction to Chapter five that there was some precedent for using eye movement measures, and particularly measures of first pass reading time and regressions as indices of referential anomaly detection. A number of researchers have used the first pass reading time to as an index of referential processing, including Ehrlich and Rayner (1983), Vonk (1984), and Garrod et al (1994), while Ehrlich and Rayner found that subjects made regressive eye movements to locate the antecedents of pronouns.

Of these studies, the current eye movement experiments had most in common with Garrod et al, who found evidence of the immediate use of information about discourse focus, gender and verb-pragmatics in reference resolution. There was evidence of first pass anomaly detection when either the gender of a pronoun or the pragmatic fit of a verb failed to match the gender or discourse role of the focal antecedent. However, the current experiments failed to replicate these findings. There was no evidence of on-line anomaly detection, i.e. a disruption of processing caused by anomalous reference to an unfocused subset that was concurrent with the reading of the anaphor. While there was some evidence of disruptions to processing during the first pass of the sentence, this was inconsistent, and so difficult to interpret as evidence of First pass anomaly detection.

There are several possible explanations for the absence of First pass reading time effects. The most obvious of these is simply that referential anomalies of the type used here are not processed on-line, possibly because an interaction of quantification and anaphora is too complex for the on-line processor to handle.

Such processing may be delayed until sufficient resources are available. This is essentially the case which Perfetti (1993) makes for minimal on-line inferential processing. Other possibilities arise from a more careful comparison with the Garrod et al study. One difference was that Garrod et al sited their anomaly manipulation on a verb, whereas the quantifier experiments used a noun. It may be that verbs play a more central role in sentence processing and produce more punctuate anomalies. This may be the most important difference between the present experiment and Garrod et al. As will be seen in a moment, the present materials were flawed by that fact that there always plausible continuations to the anaphor (*their presence/absence*) that refer to the focused subset. Also, Garrod et al used singular personal pronouns which could be unambiguously resolved by gender information, while the quantifier studies used plural possessive pronouns. These may have failed to show punctuate anomaly detection because the plural pronouns are ambiguous in their reference.

One such ambiguity is between an interpretation of the pronoun as reference to the focal subset of the quantified NP, or as reference to the superset. This means that the plural pronoun in (1) could refer to either the refset of MPs who attended the meeting, or the full set of MPs who did or did not attend the meeting. Processing of the anomaly may be delayed until one of these sets has been selected as the appropriate antecedent.

(1) A few of the MPs went to the meeting. They . . .

A second ambiguity is more subtle. Unlike singular personal pronouns, plural pronouns can refer to implicit entities (Bosch, 1988). For example, if we imagine that (2) is a question addressed to a teacher by a student, then the singular masculine pronoun requires a referent if the question is to be understood. However, the plural pronoun does not, and can instead be understood as reference to an unknown group of people who are the implicit agents of asking an exam question. In the same way, the singular masculine pronoun requires an explicit antecedent in (3), but the plural pronoun refers to an implicit or contextually determined antecedent. This might be people from Edinburgh in general, or language researchers, or whoever, depending on the context of the utterance. As before, anomaly detection may be delayed until the processor has determined the appropriate antecedent. Presumably the plural possessive pronoun *their* will exhibit the same ambiguity as *they* does.

(2) What questions will he /they ask in the exam?

(from an example suggested by Martin Pickering).

(3) He / they work(s) hard in Edinburgh. (from Yule, 1981).

This gives us two possible explanations for the absence of punctuate referential anomaly detection. It may be that a manipulation of the property description of an anaphor makes a lesser impact on processing than a manipulation of its verb-role, while the ambiguous nature of plural pronouns may delay their resolution. However, this still fails to explain why there was evidence of on-line processing difficulty in the first pass regression data. In order to do so, we must consider how these two measures can be interpreted. For instance, it may prove interesting to distinguish between the detection of an anomaly, which is perhaps signalled by regressions, and the processing consequences of that anomaly, which may be more evident in a measure of first pass reading time. Under some conditions, the anomaly might be detected early, but have delayed consequences[26]. Now consider why this might happen.

The logic of the referential manipulation used in the quantifier studies was that the property described by the anaphor (*their presence* vs. *their absence*) would be immediately interpreted as either congruent or incongruent with the discourse role filled by the focal subset. However, it is possible to imagine continuations to these anaphoric NPs which would make congruous what was an apparently incongruous property description. For example, in (4) the apparent incongruity of referring to the *presence* of those MPs who were absent from the meeting has vanished by the end of the anaphoric sentence, and it becomes clear that the anaphor is actually making felicitous reference to the compset. In the same way, (5) demonstrates that the property described by the noun need not refer to the discourse role described by the quantified sentence. In both cases an apparently anomalous reference is ruled out by subsequent text.

---

[26] Other researchers (Liversedge and Pickering, 1995) have also begun to question traditional interpretations of eye movement data, and the assumption of a close coupling between eye movements and mental processes. Liversedge and Pickering found systematic differences between three measures of First pass reading time across a set of experiment data. They compared the traditional First Pass definition (Frazier and Rayner, 1987), which is the measure used in this thesis, with a two other measures which include portions of reading time following leftward regressions out of the current region of text. The Regression Path (or Cumulative) measure includes all of the reading time from entry into the current region until a leftward movement out of it. This means that all regressive fixations are included. The Right-Bounded measure includes all fixations with the current region prior to a leftward saccade out of the region, including re-visits following regressive saccades. Liversedge and Pickering found that the latter two measures were a better index of processing for some types of experimental manipulations, including those using anomaly detection paradigms. This suggests that anomaly detection can trigger regressive saccades and increased processing to resolve the anomaly, but that this increased processing may not be recorded in the region containing the anomaly. The Liversedge and Pickering measures may be a more appropriate index of the time course of processing for the current manipulation.

(4) Few of the MPs went to the meeting.
Their presence was demanded by the chairperson.
(5) **Their presence** at another event was a poor excuse.

In terms of the experimental results, this might mean that although the processor detects a possible anomaly at the anaphor, it does not commit to this interpretation, or allow it to impact on processing, until some point downstream when the defeasible interpretation has been ruled out. Moreover, since in the present study, the text downstream of the anaphor was designed to be uninformative about reference, the processing of the anomaly may have been delayed until the end of sentence. This would explain the lack of any first pass effects and the massive disruption of processing evident in the total reading time data. This is a tentative explanation, but there are two sources of evidence in its support: one from the pragmatics literature, and another from investigations of reference processing.

The pragmatics literature (e.g. Grice, 1973, in Levinson, 1982) describes a phenomena called 'defeasibility', which is defined as the cancellability of pragmatics-based inferences (or implicatures). Examples of this include the suspension or denial[27] of conversational implicatures and presuppositions. Grice (1973) defines a conversational implicature as an implication of a sentence, which is separable from an entailment. This means that if (6) is true, then this entails the truth of (7), but only implies the truth of (8).

(6) John has three cows.
(7) John has two cows.
(8) John has three cows and no more.

Grice argues that this implicature arises from the listener's assumption that a speaker will be maximally informative. However, implicatures can be cancelled by other features of the discourse, while entailments cannot. For example, sentences (9) and (10) add *if* clauses which, in (9) suspends the implicature that John has no more than three cows, and in (10) attempts to suspend the entailment that he has two. The suspension of an implicature

[27] Levinson distinguishes between the overt denial of an implicature and its suspension (Levinson, 1982, pp115, 194-195; see also Horn, 1972). Implicatures are suspended by the addition of an *if*-clause which questions the validity of the implicature or presupposition. For example, the sentence *John didn't cheat again* has the presupposition that John had cheated, but if this is given an *if*-clause which questions that presupposition, then it disappears: *John didn't cheat again, if he ever had.*

makes no difference to the acceptability of a sentence, but any attempt to suspend an entailment results in an anomaly, as demonstrated by the unacceptability of (10).

(9) John has three cows, if not more.
(10)? John has three cows, if not two.

The defeasibility of presuppositions can be demonstrated in a similar way, which means that given a sentence like (11), this presupposes that (12) is true.

(11) Sue cried before she finished her thesis.
(12) Sue finished her thesis.

However, this can be overtly denied by changing the main verb, or suspended by the addition of an *if*-clause. So, although (11) is a valid presupposition of (12), it is not a valid presupposition of either (13) or (14).

(13) Sue died before she finished her thesis.
(14) Sue cried before she finished her thesis, if in fact she ever did?

It is also interesting to note that defeasibility means that implicatures and presuppositions can be cancelled without producing any sense of anomaly (Levinson, 1982, pp 199).

Although defeasibility is a linguistic property of complete sentences, the quantifier studies may have illustrated an on-line processing correlate which applies to sentence fragments. Because the interpretation of *their presence* as an anomalous reference to the compset in (4) is defeasible, the processor may defer any commitment to this interpretation until the end of the sentence. However, it is not simply the case that all referential processing is delayed, because the measures of First pass regressions indicate that the referential anomaly is detected on-line. It just means that this anomaly detection does not impact on processing until the processor makes a commitment to the anomalous interpretation.

This is all very speculative, but the extent to which an interpretation is defeasible does appear to characterise referential processing in general. For instance, there is evidence that pronouns are immediately resolved on the basis of gender, which is semantic, or even morpho-syntactic in nature, and not defeasible. For example, in (15) gender clearly assigns the pronoun to Sue, and the subsequent attempt to cancel this only leads to a pragmatic anomaly: that poverty is not a reason to lend money.

(15) Sue lent Bill some money because she was poor.

In contrast, manipulations of the 'implicit verb causality' (see Stevenson and Vitkovich, 1986; Garnham and Oakhill, 1985; Garnham, Oakhill and Cruttenden, 1992; these results are all discussed in Chapter 3), either fail to find on-line evidence for verb constraints on pronominal reference, or else find that the relative informativeness of the verb determines the loci of resolution. Stevenson and Vitkovich found that subjects assigned a referent to a pronoun immediately on encountering a verb when that verb is informative', and defer assignment otherwise.

Furthermore, syntactic and semantic anomalies, which are non-defeasible, produce immediate and localised effects on measures of eye movements. For example, Frazier and Rayner (1982) found evidence of localised syntactic anomaly detection at *seems* when a minimal attachment interpretation is given to (16). Minimal attachment requires that the processor postulates a minimal number of nodes as it computes the syntactic representation of a sentence, which in (16) requires that *a mile* is interpreted as the direct object of *jogs*, and not as the subject of the following clause.

(16) Since Jay always jogs a mile seems like a short distance.

More recently, Pickering and Traxler (submitted) have used the detection of semantic anomalies as an index of syntactic processing. They found punctuate anomaly detection for materials like (17) when *the yacht* is interpreted as the direct object of *editing*, and not the subject of a second clause. This occurs because of the implausibility of someone editing a yacht. Since the sentence is perfectly grammatical, this is best described as a semantic anomaly, meaning that *yacht* is a poor recipient of the act of editing.

(17) While the woman was editing the yacht sailed across the harbour.

We can also relate defeasibility to the discussion of elaborative inferences in Chapter three. It was argued there that existing taxonomies of inferences cannot determine which are made on-line, and which are off-line, and that a better distinction can be made between those inferences which are under semantic control, and others which are based on pragmatic knowledge. The latter, it was argued, are unlikely to be observed on-line, unless the there are sufficient constraints on the probability of the inference. So, McKoon and Ratcliff found that when subjects read a story about an actress falling from a tower block, they failed to infer on-line that she died, possibly because there were so many alternative outcomes. However, when Garrod et al (1990) used strong contextual constraints, they did find evidence of on-line instrumental inference.

This contrast of strong and weak contextual constraints may be another way of saying that the defeasibility of a possible inference has or has not been ruled out during processing. When the context conspires to promote an inference, then the processor may make a commitment to that inference. While this is speculative, it is also more convincing than the existing attempts to delineate on-line and off-line inferences. The Garrod et al (1994) materials may also have been designed in such a way that the processor could make an early commitment to the anomalous referential interpretation.

Conclusions

Some clear conclusions can be drawn from the reported experiments. Both the SPR and eye-tracking methodologies have demonstrated that anaphoric reference to a quantified statement is processed during normal reading, and that resolution is contingent on the form of quantification. In particular, monotone-increasing quantifiers like *a few* and *many*, focus on and enables reference to the refset, while blocking reference to the compset. In contrast, monotone-decreasing quantifiers like *few* and *not many* appear to license reference to the compset. This means that under some conditions the compset is the preferred referent of reference. However, both the refset and compset are plausible referents, therefore subjects find refset reference to be much less anomalous, and more easily recovered from.

The motivation for using measures of eye movements during reading was to describe the time course of processing. However, this was less successful, and the experiments failed to find any evidence of on-line referential processing. That is, there was no evidence that subjects found difficulty in processing reference to an unfocused subset as they read the anaphor. Neither was their any reliable evidence that reference to the unfocused subset disrupted processing during the first pass through the sentence.

This chapter has included some speculations on why no effects were observed on the first pass. These were motivated by a comparison with the study conducted by Garrod et al (1994), who did find first pass evidence of referential anomaly detection. It was argued that siting the anomaly on an intransitive verb might produce more punctuate effects, since it is less likely that the remaining sentence will reverse the referential interpretation. That is, it is less likely that the initial referential interpretation will prove to be defeasible. It was also observed that plural pronouns introduce ambiguities which may mitigate against first pass effects. I want now to consider some experimental designs which might better explore the time course of referential processing.

Proposed experiments on the time course of referential processing

The results of the experiments conducted so far can be interpreted as evidence that monotone-increasing and monotone-decreasing quantifiers focus on and enable reference to different NP subsets, and the data suggests that this may be processed on-line. Just how immediate this is remains a open question, however. The following three proposed experiments are designed to investigate the locus and time course of referential processing. The first experiment incorporates several of the suggested modifications, but still employs plural pronoun reference. The second experiment replaces the plural pronoun with a singular one, while the third experiment examines the on-line commitment to a referential interpretation given the defeasibility of these inferences.

Experiment one: The on-line processing of plural pronoun reference to the refset and compset of a quantified NP

The first of these proposed experiments will use materials like (18), where the anomaly is sited on an intransitive verb-phrase (*gambled recklessly*), and the reference type manipulation is moved to the quantified clause. This makes for easier comparisons across the four experimental conditions, because the critical regions are identical. Finally, the quantified and referential clauses are linked by the causal connective *so*. Moxey and Sanford (1987) observed that causal connectives increase the incidence of compset-focused continuations in their off-line task. Presumably this will transfer to an on-line task, and the increased reliability of compset focus following monotone-decreasing quantification will enhance the contrast between the quantificational conditions, and increase the chances of seeing on-line anomaly detection. The experiment is designed to promote immediate anomaly detection on the verb-phrase

(18) At the casino
A group of men won a lot of money on the roulette wheel.
Quantifier of the men were [*careful careless*] with their winnings
so they **gambled recklessly** until the money was gone.

Experiment two: The on-line processing of singular pronoun reference to refset and compset partitions of a quantified NP

Where the first proposed design retains the plural pronoun reference which was identified as a possible source of delay, the second proposed experiment replaces this with a singular pronoun, and uses the quantifiers *one* to focus on the refset, and *all but one* to focus on the compset (see 19a & 19b).

Again, it is predicted that any referential anomaly will be immediately detected on processing the verb-phrase, but this design rules out the referential ambiguity which is inherent in plural pronouns as the explanation of any delayed effects.

(19) At the casino
A group of men won a lot of money on the roulette wheel.
(19a) *One* of the men was [*careful* | *careless*] with his winnings
(19b) *All but one* of the men was [*careful* | *careless*] with his winnings
so **he gambled recklessly** until the money was gone.

Experiment three: Is there an early or late commitment to referential interpretations?

A third experiment is intended to determine whether the processor makes any early commitment to an anomalous referential interpretation. The design uses the same materials as the previous proposed designs, but removes the causal conjunction to produce separate quantified and referential sentences. In one condition the referential sentence begins with the adversative conjunction *nevertheless* (20a), and in other this is placed in sentence final position (19b). When the conjunction is used early in the sentence, it rules out the anomalous interpretation of the noun-phrase, but can only reverse this interpretation when it appears late in the sentence. If the processor makes an early commitment to the anomalous interpretation, then (20b) will have a longer reading time than (20a). A late commitment will produce no difference in reading time between the two sentences.

(20) At the casino
A group of men won a lot of money on the roulette wheel.
Quantifier of the men had wanted to save their winnings.

(20a) *Nevertheless* they **gambled recklessly** until the money was gone.
(20b) They **gambled recklessly** until the money was gone *nevertheless*.

Other manipulations might determine the contribution made by conjunctions. The first proposed experiment includes a causal conjunction as a means of enhancing compset focus. Other studies could compare this with an unconjoined sentence, or ones containing stronger causal conjunctions (*and so*), or additive (*and*) or adversative (*yet*) conjunctions which are seen to inhibit compset focus in off-line tasks (Moxey and Sanford, 1987).

Chapter seven:

The locus of referential anomaly detection[28].

Introduction

The study reported in this Chapter is the first of the experiments proposed in the previous chapter. It revised the design and methodology employed in Experiment 3 to measure eye movements while processing reference to the refset or compset of a sentence quantified by either *a few* or *few*. In this experiment both the quantificational and referential manipulations were made in the quantified sentence, and the referential anomaly was sited on an intransitive verb-phrase. This meant that it was possible to make a direct comparison of the reading time measures within the same quantificational condition, and reduced the plausibility of continuations that reverse the initial referential interpretation. That is, siting the anomaly on an intransitive verb-phrase increased the likelihood of observing a punctuate effect. Finally, the causal connective *so* was used to conjoin the quantified and referential sentences, since causal connective increase the frequency of compset-focused interpretations (Moxey and Sanford, 1987).

Method

Design

The experiment followed the same design as the previous SPR and eye-movement experiments. There were two within-subjects manipulations. The subject NP of the quantified sentence was quantified by either the monotone-increasing *a few* or the monotone decreasing *few*. The predicate of the anaphoric sentence was also manipulated in order that subject NP of the following sentence described a plausible action of either the refset or compset of the quantified sentence.

Subjects

Thirty-two undergraduates from the University of Glasgow were paid £5 each for their participation. As before, subjects were set-up on the eye-tracker and calibrated within a set criterion for recording accuracy prior to viewing the experiments materials. Those who either could not be set-up, or failed to make the criterion level during calibration were replaced and paid a lesser amount.

---

[28] I am grateful to Eugene Dawydiak at the University of Glasgow for his help in running this experiment, and to Simon Liversedge who provided filler materials and helped run the study.

Other subjects were replaced following data collection, but prior to analysis, if there was missing data for more than 8 experimental items (25%). Failed items included those where no data was available because of either subject or tracker loss, and items where zero fixation time was recorded for either of two large text regions (one which contained a context-setting first sentence, and another which contained the quantified sentence), or two successive regions of text. Again this was attributed to either subject or tracker error. Six subjects were replaced for these reasons.

Apparatus

As with Experiments three and four, a SRI Dual-Purkinje Eye-tracker was used to monitor eye movements, and interfaced with a 486 PC clone and button box. A bite bar and head restraint were used to minimise subjects' head movements.

Materials

Thirty-two experimental materials were used. These had a different structure from the materials used in Experiments three and four. In particular, both experimental manipulations were made in the quantified sentence, and the causal conjunction *so* was used to conjoin the quantified and anaphoric sentences. This was intended to increase the likelihood of compset focus following quantification by *few* (Moxey and Sanford, 1987). Other differences were that the plural subject pronoun *they* was used to refer to the quantified NP, and was followed by a verb-phrase which was either congruous or incongruous with the discourse role of the focal subset. The previous studies had used a possessive plural pronoun followed by a noun to describe a congruous or incongruous property of the subset.

The verb-phrases comprised an intransitive verb followed by an adverb. For half of the materials this verb-phrase was congruous with the positive version of the VP in the quantified sentence, and for the other half it was congruous with the negative version of this VP. For example, in (1) the verb *surrendered* is consonant with conceding defeat, and in (2) *gambled* is consonant with being careless with the casino winnings.

(1) After a hijacking
The hijackers were trapped in the aircraft and surrounded by police.
Quantifier of the hijackers decided to **[concede | resist]** defeat,
so they *surrendered unconditionally* before the police attacked.

(2) At the casino

A group of men won a lot of money on the roulette wheel.
Quantifier of the men were **[careful | careless]** with their winnings,
so they *gambled recklessly* until the money was gone.

The passages were always presented with the anaphoric sentence on a separate line, which meant that the critical verb always began 8 characters from the left hand edge of the text.

The experimental materials were mixed with 24 filler items which were materials for another experiment, and another 11 filler items. Three of these filler items were placed at the beginning of the experiment to orientate subjects to the task.

Subjects viewed each of the experimental passages in only one of the experimental conditions, and the materials were placed into experimental conditions following a Latin Square design. This produced four experimental blocks for each of the materials in each of the experimental conditions, and meant that there were 8 observations per cell.

The experimental materials and filler items are listed in Appendix 2.

Procedure

The experiment followed the same procedure as Experiments three and four, but with the additional use of a bite bar (in place of the chin rest) to minimise head movements during reading. This produces more accurate data, and was not available at the time of running the first two eye movement experiments.

Data Analysis

Again, this experiment followed the same analysis procedures as Experiments three and four, with a pre-analysis phase where fixation y co-ordinates were corrected by hand, before an automated procedure summed short fixations within a given range of characters. Both of these procedures are more fully described in Chapter 5.

The experimental materials were then split into regions, and Total and First pass reading time measures were calculated for these regions. The region divisions are illustrated in (3), where a slash denotes a region division. The first region is the context region, which includes the title and first sentence. The next region is the quantified sentence, and the subsequent anaphoric sentence was divided into six regions. The first contained the conjunction *so*, the next the anaphor and verb, then the adverb, temporal conjunction region, following NP and finally the VP at the end of the sentence. Gaze duration measures are also reported for a region which contains the anaphor, verb and adverb (*they gambled recklessly*).

(3) At the casino
A group of men won a lot of money on the roulette wheel.
/**Quantifier** of the men were **[careful | careless]** with their winnings,
/ so /they *gambled* /*recklessly* /until /the money /was gone.

Results

As with the previous eye-movement studies, items were removed prior to analysis if zero fixations were recorded for the context region or quantified sentence, or if there were zero fixations for two consecutive regions of the anaphoric sentence. The removed items were separately calculated prior to the Total and First pass reading time analyses. About 4% of items were lost prior to the Total Time analysis, and about 7% prior to First Pass analysis.

Total reading time

Separate 2x2 within subjects ANOVAs were used to analyse the Total reading time recorded for each region. This produced significant effects on the quantified sentence, and across all six regions of the subsequent anaphoric sentence.

Results for the quantified sentence

An analysis of the quantified sentence produced a main effect of monotonicity type ($F1(1,31)=9.21$, $p<0.005$; $F2(1,31)=7.76$, $p<0.01$), and an interaction of monotonicity and reference types ($F1(1,31)=6.60$, $p<0.016$; $F2(1,31)=4.99$, $p<0.033$). This is illustrated in Figure 1, which suggests that the effect is primarily due to a shorter reading time for refset reference following quantification by *a few* compared to the other three conditions.

This was confirmed by a simple effects analysis of the means, which found a significant difference between refset and compset reference following quantification by *a few* ($F1(1,31)=5.69$, $p<0.024$; $F2(1,31)=4.96$, p,0.034), but no comparable difference following quantification by *few* ($F1(1,31)=1.56$, $p<0.222$; $F2(1,31)=0.87$, $p<0.358$). Neither was there a difference between compset reference following quantification by *a few*, and refset reference following quantification by *few* ($F1(1,31)=2.54$, $p<0.122$; $F2(1,31)=1.65$, $p<0.209$); or any difference between compset reference following quantification by *a few* and compset reference following quantification by *few* (F1(1,31)=0.12, p<0.733; F2(1,31)=0.15, p<0.728). However, the average of these three conditions was significantly different from refset reference following quantification by *a few* (F1(1,31)=13.78, p<0.001; F2(1,31)=11.53, p<0.002).

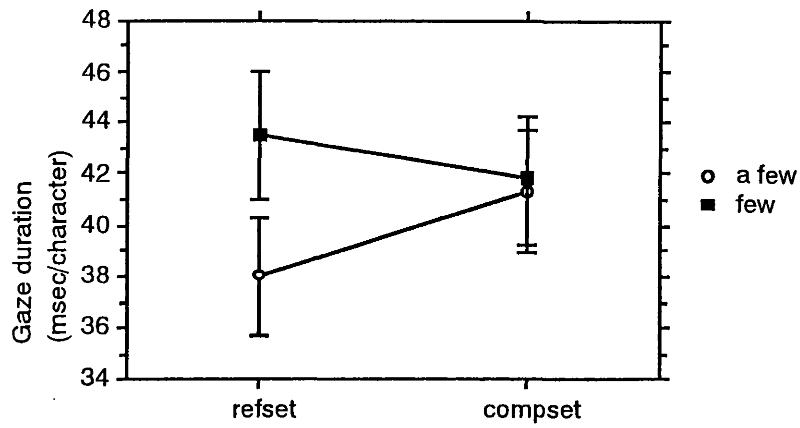


Figure 1: Mean Total reading time in msec /character (with standard error bars) for the quantified sentence following quantification by *a few* and *few*

Results for the anaphoric sentence

The mean Total reading time for the six regions of the anaphoric sentence are illustrated in Figure 2. The data points do not show the same clear divergence of felicitous and anomalous reference as the previous experiments, but suggest a more complex pattern of results. There is a consistently long reading time for compset reference following quantification by *a few*, which has the longest reading time from the anaphoric region until the end of the sentence. At the same time, refset reference to an NP quantified by *a few* shows the shortest reading time across the anaphor, adverb and temporal conjunction regions, and remains substantially shorter across the NP and VP regions compared to the reading time for reference to the refset following quantification by *few*. As with the previous experiments, this suggests that there is considerable difficulty in recovering from anomalous reference to an NP quantified by *a few*.

The reading time for the two *few* quantification conditions are similar and fall between the two *a few* conditions for as far as the NP region, where there is a marked divergence, with a longer reading time for the refset reference condition. This divergence continues into the VP region. However, the Total reading for refset reference following quantification by *few* is never as large as that for compset reference following quantification by *a few*. This suggests that reference is generally difficult following quantification by *few*, but does not produce the marked anomaly which is experienced following reference to the compset of an NP quantified by *a few*.

The analyses for the individual experimental regions are presented separately.

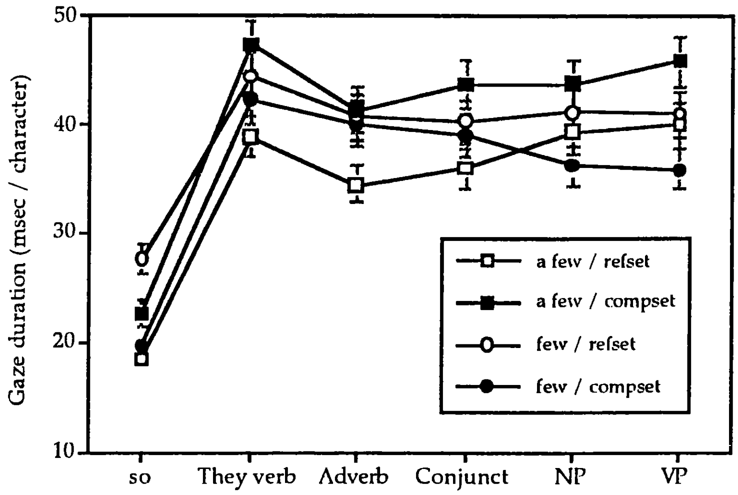


Figure 2: Mean Total reading time in msec /character (with standard error bars) across six regions divisions for the anaphoric sentence following quantification by *a few* and *few*

Conjunction region

An analysis of Total reading time in the conjunction region produced a significant main effect of monotonicity type by subjects only ($F1(1,31)=4.52$, $p<0.042$; $F2(1,31)=1.68$, $p<0.204$), and a highly significant interaction of monotonicity and reference types ($F1,31)=14.42$, $p<0.001$; $F2(1,31)=7.13$, $p<0.012$).

The mean reading times are illustrated in Figure 3, where the above effects appear to be primarily due to a difference between the reference conditions following quantification by *few*, with a markedly longer reading time following refset reference. This was confirmed by a simple effects contrast of the means, which found a significant difference between refset and compset reference following quantification by *few* ($F1(1,31)=12.41$, $p<0.002$; $F2(1,31)=7.13$, $p<0.012$); but no difference between these conditions following quantification by *a few* ($F1(1,31)=3.41$, $p<0.075$; $F2(1,31)=2.09$, $p<0.158$).

Further means comparisons established that there was no significant difference between compset reference following quantification by *few*, and refset reference following quantification by *a few* ($F1(1,31)=0.27$, $p<0.606$; $F2(1,31)=0.15$, $p<0.706$); or between compset reference following quantification by *few*, and compset reference following quantification by *a few* ($F1(1,31)=1.76$, $p<0.195$;

F2(1,31)=0.84, p<0.308). But there was a difference between these three conditions and refset reference following quantification by *few* (F1(1,31)=15.90, p<0.001; F2(1,31)=7.94, p<0.009). This suggests that the experimental effects are entirely due to an increased reading time following refset reference to an NP quantified by *few*.

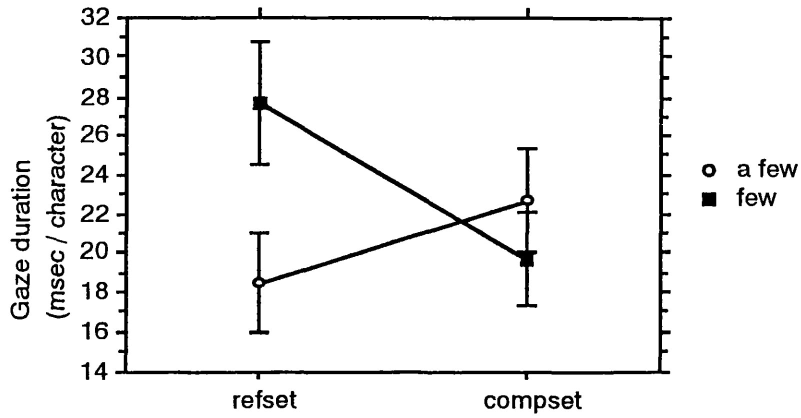


Figure 3: Mean Total reading time in msec /character (with standard error bars) for the conjunction region following quantification by *a few* and *few*

Anaphor + verb region

An analysis of Total reading time within the anaphor + verb region produced a main effect of reference type (F1(1,31)=4.54, p<0.041; F2(1,31)=5.08, p<0.032), with an overall longer reading time for reference to the compset; and an interaction of monotonicity and reference type (F1(1,31)=12.64, p<0.002; F2(1,31)=11.64, p<0.002).

The mean results are illustrated in Figure 4, and suggest that the effect is primarily caused by a difference between refset and compset reference following quantification by *a few*. This was confirmed by a simple effects analysis of the means, which established a significant difference between the reference conditions following quantification by *a few* (F1(1,31)=15.99, p<0.001; F2(1,31)=14.81, p<0.001), but no difference between these conditions following quantification by *few* (F1(1,31)=1.06, p<0.312; F2(1,31)=0.95, p<0.337). Further contrasts established that the two *few* quantification conditions had a significantly shorter reading time than that found for compset reference following quantification by *a few* (F1(1,31)=5.85, p<0.022; F2(1,31)=5.21, p<0.030); and a significantly longer reading time compared to refset reference following quantification by *a few* (F1(1,31)=4.83, p<0.036; F2(1,31)=4.67, p<0.040).

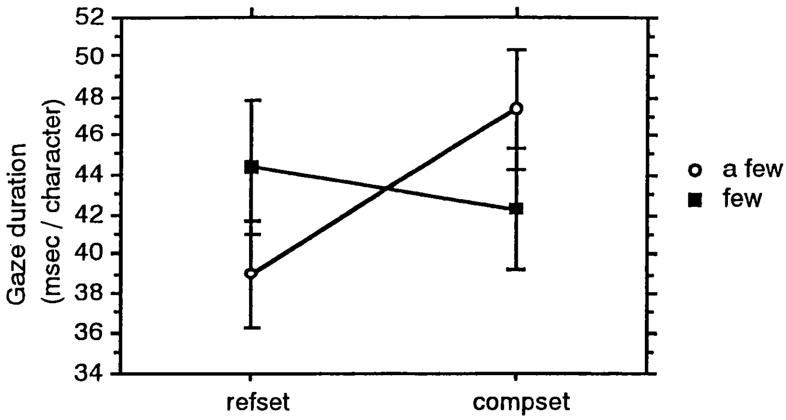


Figure 4: Mean Total reading time in msec / character (with standard error bars) for the anaphor + verb region following quantification by *a few* and *few*

Adverb region

The analysis of Total reading time for the adverb region produced an interaction of monotonicity and reference types ($F1(1,31)=4.86$, $p<0.036$; $F2(1,31)=5.54$, $p<0.026$), and an inspection of the means (illustrated in Figure 5) suggests that this effect is due to a shorter reading time for refset reference following quantification by *a few*.

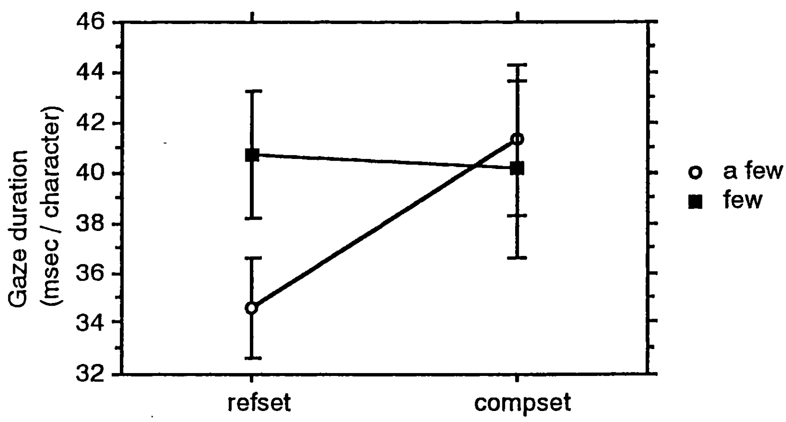


Figure 5: Mean Total reading time in msec / character (with standard error bars) for the adverb region following quantification by *a few* and *few*

This was confirmed by an analysis of the simple effects contrasts, which produced a significant difference between the reference conditions following quantification by *a few* ($F1(1,31)=8.31$, $p<0.008$; $F2(1,31)=8.82$, $p<0.006$), but no comparable difference following quantification by *few* ($F1(1,31)=0.06$, $p<0.816$; $F2(1,31)=0.13$, $p<0.722$). Moreover, there was no difference between a comparison of compset reference following quantification by *a few* and refset reference following quantification by *few* ($F1(1,31)=0.07$, $p<0.799$; $F2(1,31)=0.03$, $p<0.871$); or between compset reference to both types of quantified NPs ($F1(1,31)=0.24$, $p<0.627$; $F2(1,31)=0.28$, $p<0.604$). However, there was a significant difference between the average of these three conditions and refset reference following quantification by *a few* ($F1(1,31)=10.40$, $p<0.004$; $F2(1,31)=11.26$, $p<0.003$).

Temporal conjunction region

The analysis of Total reading time for the temporal conjunction region produced a marginal interaction ($F1(1,31)=3.33$, $p<0.078$; $F2(1,31)=3.84$, $p<0.059$). An inspection of the means (illustrated in Figure 2) suggests that the interaction was again primarily due to a difference between refset and compset reference following quantification by *a few*. However, no further analyses were conducted because of the lack of significance of the main analysis.

Noun-phrase region

The analysis of Total reading time for this region produced another interaction of monotonicity and reference type ($F1(1,31)=9.70$, $p<0.004$; $F2(1,31)=13.04$, $p<0.002$), but also a marginal main effect of monotonicity type ($F1(1,31)=4.03$, $p<0.054$; $F2(1,31)=3.70$, $p<0.012$) which is due to an overall longer reading time following quantification by *a few*. An inspection of the mean reading times (see Figure 6) suggests that the marginal main effect is caused by an overall longer reading time following quantification by *a few*; and the interaction is caused by a shorter reading time for compset reference following quantification by *few*, and a longer reading time for compset reference following quantification by *a few*.

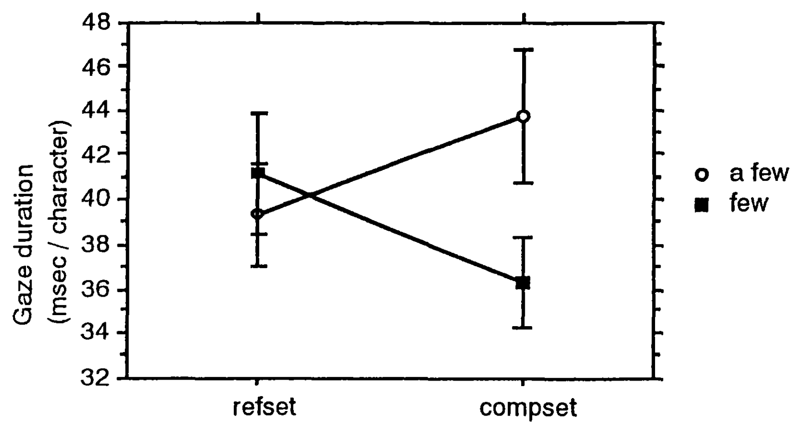


Figure 6: Mean Total reading time in msec /character (with standard error bars) for the noun-phrase region following quantification by *a few* and *few*

An simple effects analysis of the means confirmed that there is a significant difference between refset and compset reference following quantification by *a few* (F1(1,31)=4.42, $p<0.044$; F2(1,31)=5.81, $p<0.023$), and between these two reference conditions following quantification by *few* (F1(1,31)=5.31, $p<0.029$; F2(1,31)=7.27, $p<0.012$). There was also difference between compset reference for the two quantificational conditions (F1(1,31)=12.58, $p<0.002$; F2(1,31)=16.40, $p<0.001$), with a longer reading time for compset reference following quantification by *a few*. However, there was no difference between refset reference in the two quantificational conditions (F1(1,31)=0.74, $p<0.398$; F2(1,31)=1.12, $p<0.299$); between refset reference following quantification by *a few*, and compset reference following quantification by *few* (F1(1,31)=1.55, $p<0.223$; F2(1,31)=1.83, $p<0.186$); or compset reference following quantification by *a few* and refset reference following quantification by *few* (F1(1,31)=12.58, $p<0.002$; F2(1,31)=11.72, $p<0.002$).

This suggests that the interaction is due to a difference in reference type in opposite directions for both quantificational conditions, and a longer reading time for compset reference following quantification by *a few* which is countered by a shorter reading time for compset reference following quantification by *few*.

Verb-phrase region

Finally, an analysis of the Total reading time for the verb-phrase at the end of the anaphoric sentence produced similar effects to those found on the analysis of the noun-phrase. There was a main effect of monotonicity type (F1(1,31)=6.75, $p<0.015$; F2(1,31)=8.03, $p<0.009$), which is due to an overall longer reading time following quantification by *a few*; and an interaction of monotonicity and reference type (F1(1,31)=7.30, $p<0.012$; F2(1,31)=11.03, $p<0.003$). An inspection of the means (see Figure 7) suggests that this is again due to a longer reading time for compset reference to an NP quantified by *a few*, and a shorter reading time for compset reference to the same NP quantified by *few*.

A simple effects comparison of the means also found a significant difference between refset and compset reference following quantification by *a few* (F1(1,31)=4.19, $p<0.050$; F2(1,31)=7.30, $p<0.012$); but a difference between these two reference conditions following quantification by *few* which was only marginal by both subjects and items (F1(1,31)=3.15, $p<0.086$; F2(1,31)=3.98, $p<0.055$). Refset reference was not significantly different between the two quantificational conditions (F1(1,31)=0.14, $p<0.708$; F2(1,31)=0.14, $p<0.708$); nor was there a significant difference between refset reference following quantification by *a few*, and compset reference following quantification by *few*. The difference between these three conditions and compset reference following quantification by *a few* was significant ($F1(1,31)=8.55$, $p<0.007$; $F2(1,31)=14.89$, $p<0.001$). This suggests that the experimental effects are primarily due to longer reading time following anomalous reference to an NP quantified by *a few*.

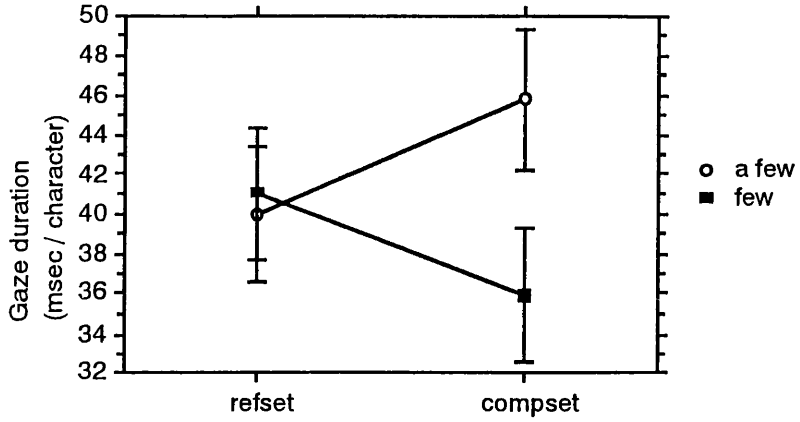


Figure 7: Mean Total reading time in msec /character (with standard error bars) for the verb-phrase region following quantification by a few and few

There were no other significant effects on any of the regions (all F's<2.1)

Summary of Total reading time results

There are three main components to the pattern of Total reading time results. The first of these is the quantified sentence, where there was a longer reading time for both the two *few* quantification conditions, and for compset reference following quantification by *a few*, compared to refset reference to an NP quantified by *a few*. If it turns out that there is no subsequent evidence of First pass differences for the quantified sentences, then these results can safely be interpreted as a description of second pass processing in response to later difficulties. This suggests that subject return to the quantified sentence following anomalous reference to an NP quantified by *a few*, and that there is a general difficulty in processing reference to one quantified by *few*. Moreover, this difficulty is as marked as anomalous reference in the monotone-increasing condition.

The second area of interest contains the anaphor+verb, adverb and temporal conjunction regions, where there is a similar differentiation of the two quantifiers. All three regions (with marginally significant results for the conjunction region) show longer Total reading times for refset compared to compset reference in the *a few* quantification condition, but with no difference between these reference conditions following quantification by *few*. The two *few* conditions have as long a reading time as that for anomalous reference to an NP quantified by *a few* at the adverb region, but lie between *a few*'s anomalous and felicitous reference conditions at the anaphor+verb and temporal conjunction regions. Again these results can be interpreted as evidence for increased processing difficulty following compset reference to a monotone-increasing quantified NP, and a general difficulty in processing reference to a monotone-decreasing quantified one.

Finally, analyses of the NP and VP regions produce a true interaction, with longer reading time for compset reference following quantification by *a few*, and for refset reference following quantification by *few*. The result for *a few* are consistent with those found in the earlier regions, but there are two possible interpretations of the data for *few*. Either refset reference leads to increased Total reading time which is localised at the NP and VP regions, or else it is the difficulty in processing compset reference which is localised at the anaphor, adverb and conjunction regions. When this disappears at the NP and VP regions, the difficulty in processing refset reference following *few* quantification which is spread across the entire sentence, just becomes more apparent. The latter explanation is most likely, given that it is consistent with observations of a difficulty in processing reference to an NP quantified by *few* which is commensurate with the difficulty of anomalous reference to an NP quantified by *a few*; and that it is not clear how the NP and VP regions would aid the interpretation of anomalous reference on their own.

The only complication comes from the results for the causal conjunction region. Here there was a markedly longer reading time following refset reference to an NP quantified by *few* compared to the other three conditions. This may indicate that subjects pay particular attention to the conjunction when processing anomalous reference to an NP quantified by *few*, which is possible, given that causal conjunctions do seem to influence the focus of monotone-decreasing quantifiers (Moxey and Sanford). However, it would be a mistake to place too much weight on the results for this region, because it is likely that any observed effects are skewed by other factors. These will include the small size of the region, and its early line position and sensitivity to errors in landing position following the return sweep from the end of the previous line. There was also a high incidence of word-skipping for this region (it was skipped on 62% of items), which may also confound the observed effects.

Anaphor + verb + adverb region

The Total reading times for a region containing the anaphor, verb and adverb was also conducted. There was a main effect of reference type ($F1(1,31)=8.30$, $p<0.008$; $F2(1,31)=6.81$, $p<0.014$), which is due to an overall longer reading time following reference to the compset, and an interaction of monotonicity and reference types ($F1(1,31)=10.61$, $p<0.003$; $F2(1,31)=10.65$, $p<0.003$).

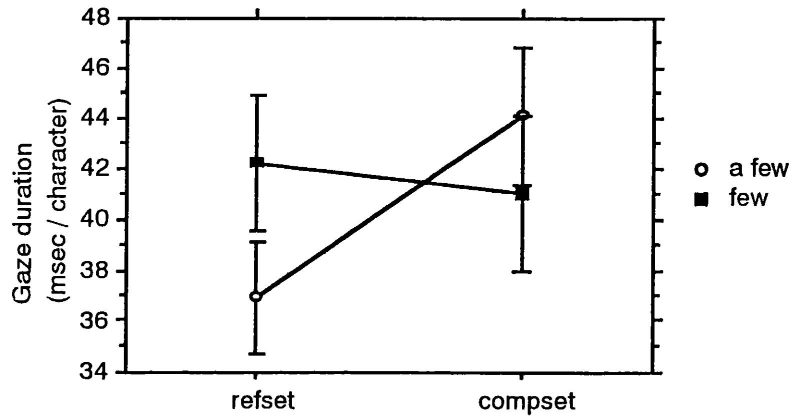


Figure 8: Mean Total reading time in msec / character (with standard error bars) for the anaphor + verb + adverb region following quantification by *a few* and *few*

Analysis of the simple effects confirmed that this interaction is entirely due to a contrast between the reference conditions following quantification by *a few* ($F1(1,31)=15.69$, $p<0.001$; $F2(1,31)=15.28$, $p<0.001$), with no difference between the same reference conditions following quantification by *few* ($F1(1,31)=0.42$, $p<0.524$; $F2(1,31)=0.48$, $p<0.494$).. Neither was there a difference between compset reference conditions across the two quantifiers ($F1(1,31)=2.84$, $p<0.103$; $F2(1,31)=2.37$, $p<0.135$), nor between compset reference following quantification by a few and refset reference following quantification by few ($F1(1,31)=1.08$, $p<0.307$; $F2(1,31)=0.72$, $p<0.405$). However there was a difference between these three conditions and refset reference following quantification by *a few* ($F1(1,31)=13.98$, $p<0.001$; $F2(1,31)=14.67$, $p<0.001$). The observed effects are principally due to a shorter reading time for this condition.

First pass reading time

As with the Total Time analysis, separate 2x2 within subjects ANOVAs were used to analyse the First pass reading time recorded for each region. There were no significant effects for any of the regions prior to the anaphor (all $Fs<1.4$). The mean reading times for the subsequent regions, from the anaphor+verb region to the end of the anaphoric sentence, are illustrated in

Figure 8. The graph suggests that there are no effects until the NP and VP regions at the end of the sentence. This was confirmed by non-significant results for separate analyses of the anaphor + verb, adverb and conjunction regions (all Fs<1.8), while there were significant results for the NP and VP regions.

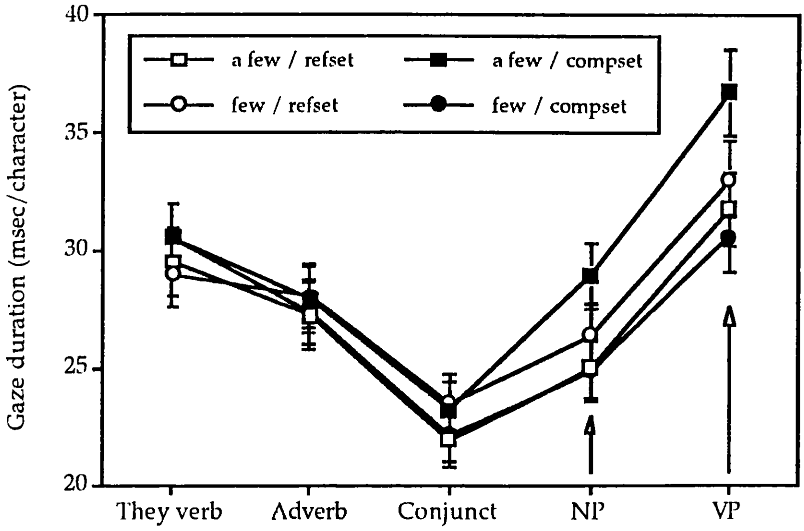


Figure 9: Mean First pass reading time in msec / character (with standard error bars) across six regions divisions for the anaphoric sentence following quantification by *a few* and *few*

Noun-phrase region

An analysis of the First pass reading time for this region produced an interaction of monotonicity and reference types which was significant by subjects and marginal by items (F1(1,31)=4.80, p<0.037; F1(1,31)=3.37, p<0.077). An inspection of these means suggests that this effect is principally due to a longer reading time for compset reference to an NP which is quantified by *a few*.

This was further supported by an analysis of the simple effects mean contrasts which showed a difference in reference type under quantification by *a few* which was significant by subjects (F1(1,31)=5.04, p<0.032; F2(1,31)=2.51, p<0.123), but no difference in reference type following quantification by *few* (F1(1,31)=0.725, p<0.402; F2(1,31)=1.02, p<0.321). Neither was there any difference between refset reference in the two quantificational conditions (F1(1,31)=0.56, p<0.460; F2(1,31)=0.44, p<0.512); or between refset reference following quantification by *a few* and compset reference following quantification by *few* (F1(1,31)=0.01, p<0.920; F2(1,31)=0.12, p<0.733). However, there was a difference between these three conditions and compset reference following quantification by *a few*, which was significant on a subjects analysis ($F1(1,31)=6.18$, $p<0.020$; $F2(1,31)=3.29$, $p<0.080$).

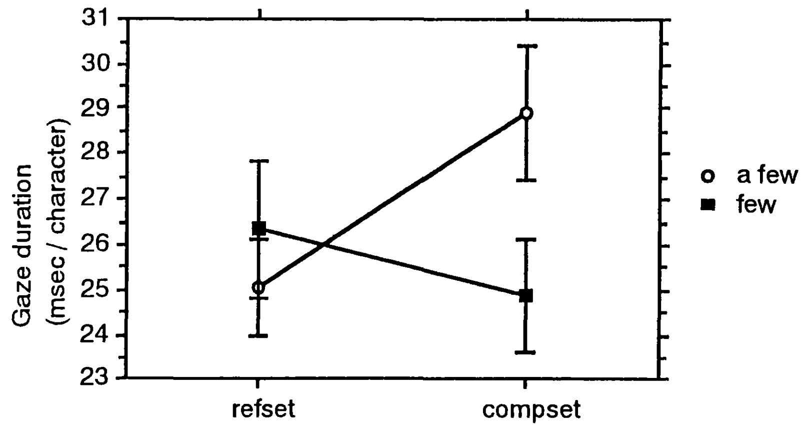


Figure 10: Mean First pass reading time in msec /character (with standard error bars) for the noun-phrase region following quantification by *a few* and *few*

Verb-phrase region

A similar pattern of results was obtained from an analysis of First pass reading time for the verb-phrase region. There was an interaction of monotonicity and reference types ($F1(1,31)=4.53$, $p<0.042$; $F2(1,31)=9.85$, $p<0.004$), which appears to be due to a longer reading time for compset reference following quantification by *a few*.

Again, this was confirmed by an analysis of the simple effects mean contrasts, which showed a significant difference between reference conditions following quantification by *a few* ($F1(1,31)=4.12$, $p<0.052$; $F2(1,31)=10.63$, $p<0.003$), but no comparable difference following quantification by *few* ($F1(1,31)=0.96$, $p<0.334$; $F2(1,31)=1.39$, $p<0.248$). There was also no difference between refset reference in the two quantificational conditions ($F1(1,31)=0.21$, $p<0.600$; $F2(1,31)=67$, $p<0.419$), or between refset reference following quantification by *a few* and compset reference following quantification by *few* ($F1(1,31)=0.20$, $p<0.656$; $F2(1,31)=0.13$, $p<0.723$). There was difference between these three conditions and compset reference to an NP quantified by *a few* ($F1(1,31)=6.01$, $p<0.021$; $F2(1,31)=14.47$, $p<0.001$).

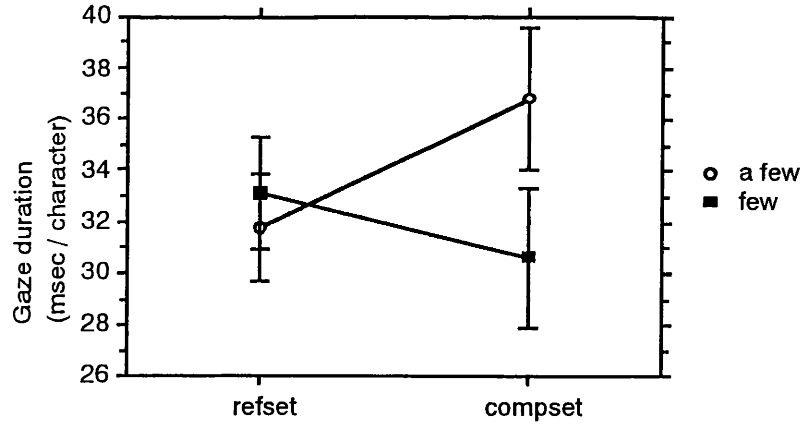


Figure 11: Mean First pass reading time in msec /character (with standard error bars) for the verb-phrase region following quantification by *a few* and *few*

The overall increase in reading time for this region compared to the previous region is probably due to sentence wrap-up processes.

Anaphor + verb + adverb region

So far the data suggests that first pass effects are delayed until some distance downstream of the anaphor. Analyses of the anaphor + verb and adverb regions both failed to identify any punctuate first pass effects. However, to be certain that such a punctuate effect was not spread across of these regions, a further analysis of First pass reading time was carried out on an combined anaphor + verb + adverb region. This produced a significant main effect of reference type ($F1(,131)=9.58$, $p<0.005$; $F2(1,31)=5.38$, $p<0.028$), with an overall shorter reading time when the anaphor is congruous with the discourse role filled by the refset. The mean First pass reading time results are illustrated in Figure 11.

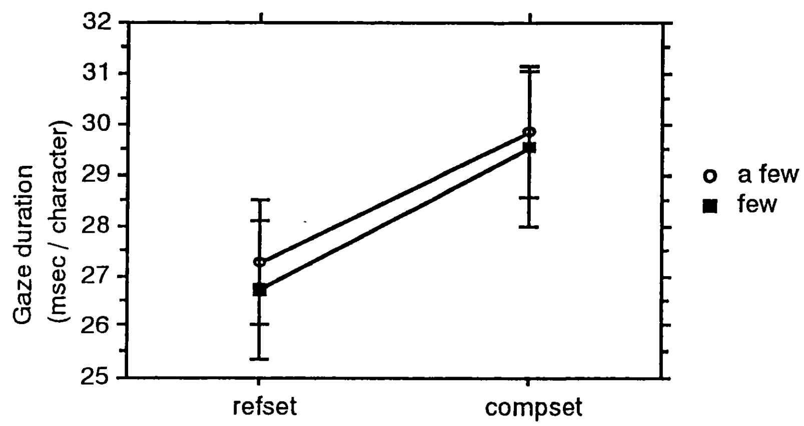


Figure 12: Mean First pass reading time in msec /character (with standard error bars) for the anaphor + verb + adverb region following quantification by *a few* and *few*

First Pass regressions
The mean first pass regressions made from each of the regions following the anaphoric manipulation (the anaphor + verb to final VP regions) were separately analysed using within subjects 2x2 ANOVAs. However, there was no evidence of a difference in frequency of regressions across the experimental conditions for any of the analysed regions (all F's<2.1).

Discussion
The results from this revised study of quantifier focus and anaphoric reference replicated the earlier finding that the ease of pronominal reference to subsets of a quantified NP is contingent on the form of quantification. At the same time, there was no strong evidence of on-line anomaly detection.

An analysis of the Total reading times produced the same evidence of an increased reading time for reference to the unfocused subset of the quantified NP as found in the previous experiments, but with a more idiosyncratic pattern of results across the individual quantificational conditions. As predicted, there was a longer reading time when the anaphoric VP described a property of the compset following quantification by *a few* than when it described a property of the refset partition. This was observed at the quantified sentence, and across all regions of the anaphoric sentence, and is consistent with the claim that monotone-increasing quantifiers like *a few* focus processing attention onto the refset, and block compset reference. At the same time, there was evidence of a more general difficulty in processing reference following quantification by *few*, with no difference in reading time for refset or compset reference at either the quantified sentence, or anaphor+verb, adverb or temporal conjunction regions of the anaphoric sentence. These reading times were often as long as those for anomalous reference to the compset following quantification by *a few*, and always longer than reading times in the contrasting refset reference condition. However, the results for other regions suggest that there was still a preference for compset reference following this type of quantification. The reading times for the causal conjunction, NP and VP regions of the anaphoric sentence were significantly longer in the refset reference condition.

In summary, this pattern of results support the earlier conclusion that monotone-increasing quantification focuses on and enables reference to the refset but blocks reference to the compset, to such an extent that it seems likely that the compset is unrepresented. In contrast, monotone-decreasing quantifiers like *few* license reference to the compset, which means that both the refset and compset are plausible antecedents, but that the compset is often favoured.

According to the analysis of First pass reading times, there was no evidence of on-line anomaly detection for two quantificational conditions. However, there was evidence that reference to the compset resulted in a general increase in first pass reading time for a region of text containing the pronoun and intransitive verb-phrase. This suggested that there was a general difficulty in processing reference to the compset, whatever the prior form of quantification. These results can be interpreted in two different ways. It may indicate that quantification by *a few* resulted in focus on the refset, therefore reference to this subset was more easy to process, and so had a shorter reading time that reference to the compset. This was consistent with the experimental predictions. However, there is an alternative account, according to which it may be that a simple co-referential mapping was established between the act described by the anaphor and by the VP of the quantified sentence, regardless of quantificational information. The present results cannot separate these two possibilities.

The other First pass reading results were more easily interpreted as evidence that anomalous reference to an unfocused subset had processing consequences. There was an increased First pass reading time for both the sentence-final NP and VP regions of the anaphoric sentence when reference was made to the unfocused subset following quantification by *a few*; but there was no commensurate effect for reference to the refset following quantification by *few*. The results suggest that although there was no strong evidence of first pass anomaly detection, there is evidence that the anomaly is detected more quickly, and has an earlier impact on processing, following monotone-increasing quantification by *a few*. This may be consistent with a claim made by Dowty (unpublished) that monotone-decreasing quantification is non-referential, and that it is necessary to use pragmatics and situation-specific knowledge to resolve anaphora under these conditions.

In the light of the claims made in Chapter 6, it appears that the revised experimental design which, in particular, sited the referential anomaly on an intransitive VP was unable to provide convincing evidence of first pass anomaly detection following reference to the unfocused subset of a quantified sentence. However, the experiment still fails to address the possible ambiguities associated with plural pronominal evidence. It is still possible that punctuate referential anomaly detection will be detected by the other experiments proposed in Chapter 6.

A final point concerns the failure to find any evidence of a difference in regressions across the experimental conditions, despite finding effects in the previous experiments. There is no clear explanation for this, although it may indicate the unreliability of the previous results. Alternatively, it may be due to changing the site of the anomaly. It was sited on a noun-phrase in the previous experiments, and regressions were made either from this region, or from regions later in the sentence. However, in this experiment the anomaly was sited on a verb-phrase, which was more salient than the previous manipulation, and so may have reduced the need for eye movements to either check the content of the anaphor or the quantified sentence. Other explanations, such as it is unlikely that subjects will make regressions because the quantified sentence appeared in the previous line, are implausible since this was also true of the previous studies.

Chapter eight:

Pragmatic constraints on focus following quantification by *only a few*

Introduction

This chapter contains the first of two digressions. Up to now the experiments have addressed the semantic constraints exerted on focus by monotone-increasing and monotone-decreasing quantifiers, and the interpretation of subsequent anaphoric reference. However, *only a few* was included in the introductory chapters as an example of a quantifier which exhibits neither monotone-increasing nor monotone-decreasing properties. It proved difficult to classify on a number of semantic tests, and also produced a more complex pattern of results in a sentence-completion task (Moxey and Sanford, 1987). Under some conditions it favours focus on the refset partition of the quantified NP, but will also focus on the compset under other conditions.

The chapter reports two experimental studies on discourse focus following quantification by *only a few*. These studies explore the hypothesis that *only a few* exerts weak focusing preference which are modulated by contextual factors. These will be introduced by a recap on the formal semantic properties of *only a few*, some of which indicate the context-dependency of its interpretation; and review some sentence-continuation data reported by Moxey and Sanford (1987, 1993a). The rationale of the current manipulation will then be outlined.

Formal properties of only a few

It was shown in chapter one that *only a few* violates extensionality and conservativity constraints, which together imply a context-dependent function. Its non-extensionality is demonstrated by the possible falseness of (1) given the truth of (2), because the truth of (1) depends on what the set of surviving sailors is compared against.

(1) Only a few sailors survived.
(2) Only a few sailors were surviving sailors.

If it is compared against the total number of survivors, whether or not they are sailors, or against the total number of both survivors and non-survivors, again including those who are not sailors, then (1) may be false while (2) is true. For example, it could be that (1) is used to assert that *only a few* sailors survived relative to a greater number of passengers, or that following a large number of casualties, the only survivors were a few sailors. However, if the comparison is restricted to the set of sailors, as it is in (2), then the two sentences will be synonymous and share truth-values.

At the same time a violation of the conservativity universal is demonstrated by an interpretation of (4) where the truth of the sentence is not solely determined by that part of the VP denotation which contains the NP. This means that the truth of (3) can be established by simply inspecting the set of fishermen, for which it should be possible to say that *all* of the set survived. This holds for one reading of (4), but on another it can be understood to imply that other non-fishermen did not survive. Given this interpretation, (4) is falsified if there are surviving non-fishermen.

(3) All of the fishermen survived.
(4) Only a few of the fishermen survived.

Features of the discourse context also seem to determine whether or not *only a few* take a conservative reading. If (5) is read as a context for (6) and (7), then (6) results in the same ambiguity between non-conservative and conservative readings as described above. It can be understood to either mean that many of the fishing boats and all of the ferries sank (a non-conservative reading), or that many of the fishing boats sank while implying nothing about the ferries (the conservative reading). However, when (5) is read as the context for (7), the non-conservative reading is blocked, and the sentence only implies that many boats sunk[29].

(5) Both fishing boats and ferries were endangered by the storm.
(6) Only a few fishing boats survived.
(7) Only a few boats survived.

---

[29] Given an appropriate context, non-conservative readings are also possible for sentences quantified by other quantifiers, like *all*, which are more usually classed as conservative (cf Barwise and Cooper, 1981). Also, the availability of non-conservative readings can be promoted by stress patterns. Placing stress on *fishermen* in (4), or *fishing* in (6), emphasises a contrast with the sets non-fishermen and boats other than fishing ones respectively (cf Klein, 1994). It is possible that different patterns of stress, and using pauses break up the composite quantifier, will allow different syntactic analyses in which the quantifiers *only* and *a few* are separated. The different syntactic analyses assigned to the phrase may well account for the different interpretations.

If, as Westerståhl (1984; reviewed in Chapter 1) claims, the alternative readings are determined by the selection of different context sets for the quantified NP, then there is an obvious question of whether there are psycholinguistic correlates of the procedures needed to identify the appropriate context set. Such processes would be similar to those required for anaphoric reference by definite NPs. In fact, there would be a case for claiming some uniformity between these two sets of processes, because quantifiers belong to the class of determiners according to generalised quantifier theory (cf. Barwise and Cooper, 1981; Westerståhl, 1989).

More generally, *only a few* can be seen to have the value-judgmental function which Keenan and Stavi (1986) use as a justification for excluding *many* and *few* from their semantic analysis of natural language quantifiers. Not only is its interpretation dependent on aspects of the situation (the context set on which it operates), but it also signals that an expectation is held by the speaker. For example, when the set of surviving sailors is *only a few* relative to the set of sailors, the speaker is signalling that she expected there to be more who survived. The same is true when the comparison is made against the total set of survivors (including non-sailors), or the combined set of survivors and non-survivors. This value-judgmental function is particularly important for the current experimental manipulation, where it is predicted that the value-judgements licensed by the quantifier in a given context are used to guide focus.

The non-conservative and non-extensional readings for *only a few* are shared with most, and perhaps all of the five quantifiers considered in Chapter one. There it seemed clear that *many, not many* and *few* produce such reading, and it is probable that *a few* is both non-conservative and non-extensional, although judgements are more difficult for this quantifier. However, *only a few* does not find the same uniformity with these quantifiers on other semantic tests. For instance, it is more difficult to make intuitive judgements about its acceptability in test frames of upwards and downwards monotonicity (see 8 and 9), and those researchers who have commented its function generally categorise *only a few* as non-monotonic (Moxey and Sanford, 1987; Klein, 1994). It also difficult to judge on semantic tests of negativity and persistence.

(8) If *only a few* of the students passed the exam with ease,
then *only a few* of the students passed the exam.

(9) If *only a few* of the students passed the exam,
then *only a few* of the students passed the exam with ease.

Klein (1994) proposes that *only a few,* and also *only few,* are complex quantifiers which combine the semantic properties of the simple quantifiers *only* and *a few* or *few.* This follows from Barwise and Cooper's (1981) claim that all simple quantifiers are either monotone-increasing or decreasing, and therefore non-monotonic ones must be complex expressions which are composed of two or more simple quantifiers. Klein argues that a conjunction of the downwards monotone *only* and the upwards monotone *a few* produces the non-monotonicity of *only a few,* while a combination of *only* and the downwards monotone *few* remains monotone-decreasing.

Only a few and focus

If *only a few* is non-monotonic, then this suggests that it may not exhibit the same semantic function of either monotone-increasing or monotone-decreasing quantifiers. Moxey and Sanford (1987) included it in their sentence-continuation experiment and found a pattern of results which was quite different from that for either the monotone-increasing *a few* and *many,* or monotone-decreasing *few* and *not many.* When *only a few* was substituted into sentences like (10), subjects used their continuations to refer to the refset as often as when the sentence was quantified by *a few.* The same was true when the quantified and anaphoric sentences were conjoined using *and* or *but,* but when they were conjoined by *because* the pattern of results was more similar to that found following quantification by *few.*

(10) *Quantifier* of the MPs attended the meeting. They . . .

Moxey and Sanford argued that *because* is a compset-supporting operator, because it also increased the proportion of compset-focused continuation following quantification by *few* or *not many.* They suggest that although *only a few* appears to have a default for refset focus, it can focus on the compset when it is in the environment of a compset supporting operator like *because.* This may also be true of other situations, such as when a negative polarity item appears in the sentence. It is also possible that pragmatics will influence focus for *only a few.* For instance, Moxey and Sanford (1987) observed that while two thirds of the continuations for (11) referred to the compset, hardly any did for (12).

(11) Only a few of the children ate their ice-cream. They . . .

(12) Only a few of the MPs were at the meeting. They . . .

This might be explained by a difference in expectations associated with the two scenarios. It would normally be expected that children enjoy ice-cream, but this is violated by the assertion, in (11), that *only a few* actually ate it. Readers respond to this violation by focusing on the compset of children who did not eat the ice-cream and infer reasons about why this might have happened. For example, a plausible continuation might explain that the children had already had a heavy meal and could not eat any more. In contrast, (20) does not violate any expectations about the situation it describes (possibly because readers do not have strong expectations about MPs attending meetings), and does not result in the same frequency of compset focus. Instead focus is maintained on the refset of MPs who did go to the meeting.

While such an explanation is consistent with the value-judgement function ascribed to some quantifiers by Keenan and Stavi, the contrast between these two sentences is not well-controlled, and it is possible that other differences between the two situations are responsible for the results. The following experiments make a more detailed study of pragmatic constraints on focus, and use more carefully controlled materials. The studies manipulate the verb-phrase of a sentence quantified by *only a few* to produce two passages which describe the same situation, but make quantified assertions which either match or violate expectations about that situation. Sentences (13) and (14) illustrate the manipulation.

(13) Only a few of the women were over four feet tall.
(14) Only a few of the women were over six feet tall.

Given that the normal (UK) height for women is somewhere between five and six feet, (13) marks the small refset as unusual relative to that norm, and (14) marks a small refset which fits the norm. The reference patterns of plural pronoun continuations should reflect these differences. Sentence (13) is predicted to result in a focal refset, while (14) should result in a focal compset. By this account the function of *only a few* is simplified to that of a quantifier which marks the refset partition of the NP as small, but the significance of this is determined relative to knowledge about the situation.

None of these observations, however, explain the tendency towards refset focus in the unconjoined, *and* and *but* sentence-continuation conditions. This seems to contradict any intuitions about the non-monotonicity, or even downwards monotonicity, of *only a few*. Yet it may be better explained by the primacy of the refset representation. I have argued that only the refset is necessary representation, and that the compset is represented relative to this subset when required by monotone-decreasing quantification. Because *only a few* has no strong focal preferences of its own, the representational system defaults to refset focus (and no representation of the compset). However, compset-supporting environments, such as conjunction by *because*, switch focus to the compset. It is also possible that other situations, such as conjunction by *but*, or a pragmatically marked refset, will make *only a few* focus on the refset, although this may be indistinguishable from the default situation.

Experiment six:

Pragmatic constraints on focus for *only a few*.

Introduction

This study tested the claim that the pragmatic interpretation of a sentence quantified by *only a few* can influence focus. This depends on the quantifier having the relatively simple semantic function of marking a set as small, and allowing the pragmatics of the situation to determine the significance of this. For example, in both (13) and (14) *only a few* marked the refset of women who are respectively over four and six feet tall as small in number. The former of these two cases conforms with the expectation that most women are between five and six feet tall, and that those who are taller are a small and exceptional group. The latter case violates the same expectation by suggesting that only a small number of the women fall within the normal range. It was predicted that norm-matching sentences will maintain focus on the refset, but that norm-violation increases the tendency towards compset focus.

The following experiment uses a sentence-continuation methodology to test this claim. The experiment also includes the conjunction conditions used by Moxey and Sanford (1987) to determine whether there is evidence of an interaction between semantic and pragmatic constraints on focus. In particular, I was keen to assess the contribution of a compset-supporting operator like *because*, since causal conjunctions support compset focus. This appears to be due to the fact that many compset-focused continuations tend to be explanatory, and account for the small refset (Moxey and Sanford, 1987).

Method

The experiment used the same sentence-continuation method described by Moxey and Sanford (1987). Subjects viewed single passages and were asked to provide one or two sentence continuations to plural noun pronouns. The results were then categorised by the set described by the continuation.

Design

There were two between-subjects manipulations. A sentence quantified by *only a few* described a situation which either matched or violated norm expectations. These quantified sentences were followed by a plural pronoun (*they*), which either began a new sentence, or was connected to the quantified sentence by either *and, but* or *because.*

Each subject saw only one material in one of the experimental conditions to produce a fully between subjects design.

Subjects

600 under-graduate students at Glasgow University, who were not paid for their participation.

Materials

Five experimental materials were used, and are listed in Appendix 3.
One of the experimental materials is illustrated in (15).

(15) In a restaurant
The office party went to an up-market Italian restaurant for their Christmas dinner. Only a few of the office staff **[offered to buy the waiter a drink / ate with their mouths closed]**
. They...
and they...
but they...
because they...

Procedure

Subjects were given a sheet of paper with one of the materials in one of the experimental conditions and instructed to "Please complete the following short passage with one or two sentences of your own". Following data collection, the resulting continuations were coded by the type of reference made: 'refset', 'compset', 'superset' or 'other'.

Results

The coded continuation results are summarised in Tables 1 and 2.

An analysis of variance by chi-square (Winer, 1991) for compset continuations produced significant main effects for the VP manipulation (χ^2=24.80, df=2, p<0.01) and for conjunction type (χ^2=26.22, df=3, p<0.01). There were no significant differences between the experimental passages, and no significant interactions. The main effect of continuation type was further investigated using an analysis of residuals (Seigel and Castellan, 1988). According to this analysis, there was no significant difference from chance for the frequency of compset-focused continuations for the no conjunction condition (Residual = 0.27, p > 0.05), or the *and* (Residual = 1.1, p > 0.05) and *but* (Residual = 0.6, p > 0.05) conjunction conditions. However, the frequency of compset-focused continuations for the *because* conjunction was significantly different from chance (Residual = 1.96, p < 0.05). The mean compset continuations collapsed across the five passages are graphed in Figure 1 below.

Conjunction	refset	compset	superset	other
.	9.60	3.60	1.40	0.60
and	11.20	1.80	0.60	1.40
but	7.20	3.60	3.60	0.60
because	5.20	9.00	0.40	0.40

Table 1: mean continuation categories for refset-predicting contexts.

Conjunction	refset	compset	superset	other
.	3.80	8.40	1.20	1.60
and	5.20	7.20	1.20	1.40
but	6.00	7.20	0.40	1.40
because	3.40	11.00	0.00	0.60

Table 2: mean continuation categories for compset-predicting contexts.

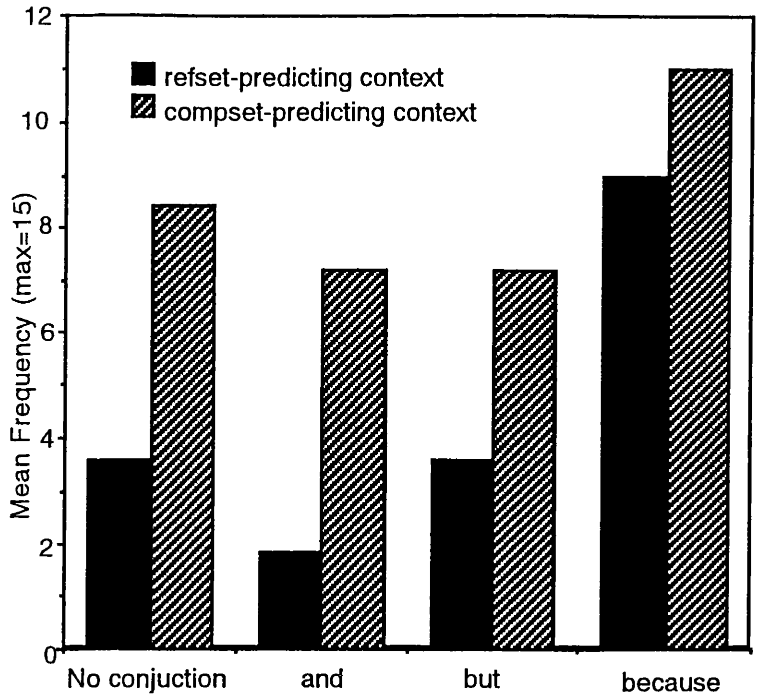


Figure 1: mean compset continuations collapsed across items following quantification by only a few in norm-violating and norm-matching pragmatic conditions.

Discussion

The first point to be made is that there was no significant difference between the five experimental materials. They worked in a sufficiently similar way to produce a difference in the proportion of compset continuation between the context types and between the connective types, and with no evidence of an interaction.

Subjects produced significantly more compset-focused continuations in the norm-violating conditions than in the norm-matching conditions, as was expected, which supports the hypothesis that focus is pragmatically mediated following quantification by *only a few*. According to the present theory, the quantifier is simply used to mark a set as small, and the significance of this is determined pragmatically. When the smallness of this set violates norm expectation, this results in an increased likelihood of focus on the compset, and defaults to refset focus otherwise.

At the same time the proportion of compset-focused continuations was increased in both contextual conditions when *because* was used to conjoin the quantified and anaphoric clauses. Conjunction by either *and* or *but* did not seem to influence the likelihood of compset focus however. This is consistent with the claim that *only a few* is likely to focus on the compset in environments containing compset-supporting operators like *because*. This may be because these operators mark the fact that the situation described by the quantifier is unusual and requires explanation.

It should be noted, however, that both of these effects only lead to increased focus on the compset, but still permitted reference to the refset. Even a combination of those conditions which favoured compset focus, where there was conjunction by *because* in a norm-violating context, only resulted in a 73% incidence of compset focus. This forces the conclusion, as similar data did for monotone-decreasing quantification, that the manipulations make the compset more focal, but cannot push the refset out of focus.

One final observation is that there was a substantially higher frequency of compset focus in the no conjunction norm-matching condition that found by Moxey and Sanford (1987) - 24% compared to 5% - but no difference in frequency between the two experiments following conjunction by because (both were 60%). This may have been because Moxey and Sanford did not control for contextual constraints in their study, and their materials described situations which were open to a number of interpretations. For example, (16) could describe a situation where many football fans were expected, say a cup final, or one where few were expected, such as a local amateur match.

(16) Only a few football fans went to the match.

Under these conditions, focus will depend on the base-rate expectations for the interpretation which subjects assign to the quantified sentence. If this is high then there will be a greater tendency towards compset focus, but if it is low this tendency will be reduced. The *because* conjunction condition may be insensitive to this just because it signals that the situation requires explanation, which indicates to subjects that a low base-rate interpretation is appropriate. In contrast, the present materials used a title and contextual sentence which narrowed the number of possible situations that the sentences could describe.

Experiment seven:

The time course of anaphoric reference to an NP quantified by *only a few*

Introduction

Having demonstrated in an off-line production task that focus is mediated by pragmatics following quantification by *only a few,* the next step is to test the same manipulation on-line and in a comprehension task. This follows the rationale that was behind the SPR and eye movement experiments reported in Chapters four and five. Moxey and Sanford had demonstrated that monotone-increasing and monotone-decreasing quantifiers produced different patterns of focus in an off-line sentence-continuation task, so the SPR and eye movement experiments were used to determine whether this also occurred during comprehension. It was possible that the effects observed by Moxey and Sanford were peculiar to production, or even particular to the task they used, and dependent on problem-solving processes.

The present study followed the design and procedure used in the previous eye movement studies, and measured the time course of anaphoric reference to refset and compset partitions of a sentence quantified by *only a few.* An example material is given in (17). As before, the verb-phrase of the quantified sentence was manipulated to describe either a norm-matching (*heaviest*) or norm-violating (*lightest*) situation. It was predicted that the norm-matching manipulation would leave the refset in focus, but that norm-violation would result in focus on the compset. At the same time, the anaphoric NP which begins the following sentence either described a property of the refset (*their strength*) or compset (*their weakness*) of the quantified NP.

(17) At the gym.
Some expert weight-lifters from the gym held a weight-lifting competition. Only a few of the weight-lifters managed to lift the [**heaviest / lightest**] dumbbell. Their [**strength / weakness**] was a surprise to their friends.

It was predicted that reference to the unfocused subset would prove anomalous and result in increased processing at or beyond the anaphor. Given the previous experimental results this might not be evident as a punctuate effect on measures of first pass reading time or regressions, but should still show as a disruption to processing on a measure of total reading time.

Before conducting this experiment, an off-line pilot study was carried out to generate a materials set.

Pilot study

The proposed experimental materials took the same format as those used in the previous eye movement experiments. They had a title and three sentences, the first of which was used to set the context. This was following by a quantified sentence of the form ***Quantifier*** **NP VP**, where the VP was manipulated to produce a predicate which either matched or violated norms for the described situation. In the norm-matching condition, the refset partition of the quantified NP was marked as small and exceptional for the given situation, while in the norm-violating situation the compset was marked as unusually large. The final sentence of each passage began with an anaphoric NP composed of the possessive plural pronoun *their*, and an NP which described a property of either the refset or compset of the earlier quantified NP.

The procedure used in the study was different in design to the one used to select materials in Chapter 4 . In that study, subjects saw both possible continuation sentences, and had to select the one which was most appropriate. This looked to be too difficult a task for the present manipulation, so subjects were instead given a sentence-continuation task, where they saw a number of proposed materials without the anaphoric sentence and were asked to produce a continuation to the plural pronoun *they*.

It was originally intended to select materials which passed a threshold of 70% refset-focused continuations in the norm-matching condition, and 70% compset-focused continuations in the norm-violating condition, but this criterion had to be relaxed when it proved difficult to generate materials with a high frequency of compset-focused continuations. Materials were finally selected if the frequency of compset-focused continuations exceeded 50% of those which could be classified as either refset- or compset-focused. This excluded those continuations which could not be classified in this way. Twenty-two materials were selected using this criterion, and an additional two were added to the material set prior to the eye-tracking study.

Fifteen continuations were collected for each verb-phrase conditions of the 22 materials and categorised by the set which they described. Subjects produced a mean 11.00 (sd=2.52) refset continuations and a mean 1.73 (sd=1.55) compset continuations for materials in the norm-matching condition. In the norm-violating condition they produced a mean 7.09 (sd=3.78) refset continuations and a mean 6.18 (sd=3.32) compset continuations. Other continuations either referred to the superset or could not be classified.

There was some concern over the failure to produce materials with a high incidence of compset focus in the norm-violating condition, but this was mitigated by the fact that the Experiment six found only a 56% incidence of compset focus in the norm-violation condition when the quantified and anaphoric sentences were unconjoined. This reflects the fact that the manipulations only increase the likelihood of compset focus, but do not push the refset out of focus.

Method

This study replicated the method of the previous eye-tracking studies reported in chapter five. The following are those details which are particular to the present study.

Design

There were two within-subject variables. The VP of the quantified sentence was manipulated to describe a situation which either matched or violated norm expectation. This was crossed with the second variable, which manipulated the anaphoric subject NP of the following sentence to describe either the refset of compset partition of the quantified sentence.

Subjects

The experiment used 24 subjects with uncorrected vision who were undergraduate students at Glasgow University and paid £5 for their participation. Subjects who failed to complete the task were paid a lesser rate. No subjects had been used in the previous experiments.

Subjects who could not be set-up or failed to calibrate on the eyetracker were replaced before viewing the experimental materials, and no subjects had to be replaced following data collection.

Materials

The experiments used 24 passages containing a sentence with a noun-phrase quantified by *only a few* and predicated by a verb-phrase describing either a norm-matching or norm-violating role for the refset. The following sentence began with an anaphoric NP which described a property of either the refset or compset partition of the quantified NP. The length of this noun-phrase was balanced across the two anaphoric conditions. It was a mean 8.67 (sd=2.10) characters in the refset-referring condition, and a mean 9.13 (sd=1.98) characters in the compset-referring condition.

The monitor used to present the materials restricted line length to a maximum 65 characters. The materials were arranged to accommodate this constraint while keeping the critical anaphoric noun-phrase of the text in an approximately central line position to avoid landing errors associated with the return sweep from the end of the preceding line (Rayner and Pollatsek, 1989). The pronoun began a minimum 8 (and a mean 18.17, sd=8.29) characters from the start of the line.

A double line-spaced presentation was used to make data correction easier during pre-analysis.

The materials were mixed with 32 filler items which were materials for other experiments, and divided into 4 presentation blocks with rest intervals between each block. Two practise items began the first block, and a further practise item was added to the beginning of the other blocks to orient subjects to the reading task. Comprehension questions requiring a YES or NO button press response followed 25% of the presented passages to ensure careful reading. Experimental materials, fillers and practise items are listed in Appendix 3.

Procedure

We followed the experimental procedure described for the previous eye-tracking experiments (see Chapter 5), with the addition of an off-line task following completion of the eye-tracking task. Subjects were given printed versions of the experimental materials which they had just viewed (excluding filler and practise items), but with both versions of the anaphoric reference placed in brackets within the text. Subjects were asked to indicate which of these two best-fitted the passage. This allowed subsequent comparison between on-line and off-line results, and provided independent criteria for subsequent data analysis.

Results

An initial analysis examined the sentence reading times. One of the clearest findings of the previous experiment was that anomalous reference produces a massive increase in reading time for the anaphoric sentence. A similar finding would demonstrate the effectiveness of the current experimental manipulation, and justify further analyses to determine the locus and time course of this processing. A corollary of this is that a failure to find sentence reading time effects would dispense with the need for these more fine-grained analyses.

The data was inspected prior to analysis, and trials were removed if zero scores were recorded for any of the sentence regions. These are due to tracker loss, or subjects not fixating within the region. Approximately 6% of the data was lost for this reason.

Sentence reading time

The reading times for the contextual, quantified and anaphoric sentences were separately analysed using 2x2 within subject ANOVA's. Only an analysis of the quantified sentence produced any significant differences which were due to the experimental manipulations, with a main effect of reference type which was significant by subjects only ($F1(1,23)=5.32$, $p<0.031$; $F2(1,23)=1.80$, $p<0.193$), and reflected a longer reading time for reference to the refset partition of the quantified NP. There were no other significant effects (all $Fs<2.0$).

Total reading time for regions of the anaphoric sentence

Although no effects were observed on the analysis of Total reading times for the anaphoric sentence, it was possible that there were short-lived effects within individual regions of the sentence. For this reason, the Total reading times were analysed for three regions of the anaphoric sentence. These regions are illustrated in (18), where the region divisions are denoted by slashes. The first region contained the anaphor, the second region contained the following verb-phrase, and the final sentence contained a prepositional phrase.

(18) Their [strength | weakness] / was a surprise / to their friends.

The Total reading times for these regions were analysed using separate 2 x 2 within subjects ANOVAs. There was no evidence of a significant effect for either the anaphor region ($Fs < 1.0$), or the verb-phrase region ($Fs < 1.0$). The analysis of Total reading time for the prepositional phrase region found no evidence of a main effect of norm-matching or violation ($F2 < 1.0$), no effect of reference type ($F1(1,23) = 1.80$, $p > 0.05$; $F2 < 1.0$), nor an interaction of these factors. These results demonstrated that there were no short-lived effects within regions of the anaphoric sentence.

First pass reading time for regions of the anaphoric sentence

The First pass reading times for the three regions of the anaphoric sentence (e.g. 18) were also analyse d using separate 2 x 2 within subjects ANOVAs. There was no evidence of any significant effects for either the anaphor region (Fs < 1.0). Nor did an analysis of the First pass reading times for the following verb-phrase region produce either a significant main effect of norm-matching or violation (Fs < 1.0), an effect of reference type ($F1(1,23) = 3.00$, $p > 0.05$; $F2 < 1.0$), or an interaction of these factors (Fs < 1.0). Finally, an analysis of the final prepositional phrase region failed to find any significant effects (Fs < 1.0). Given these results, there was no evidence that the experimental manipulations affected the First pass reading time for any regions of the anaphoric sentence.

Discussion

There was no evidence of significant effects according to either measures of the Total sentence reading time, the reading times for individual regions of the anaphoric sentence, or according to measures of the First pass reading time of the anaphoric sentence. These results allowed us to conclude that the experimental manipulations failed to affect the processing of anaphoric reference to the refset and compset of a sentence quantified by *only a few*. This meant that not only was there no evidence that norm-violation resulted in focus on the compset, but that there was no evidence of a difference between reference to either the refset and compset when the quantified sentence described a situation which was consistent with norm-expectations, and therefore was predicted to maintain focus on the refset.

The one finding which was significant (by subjects only) according to the measure of Total reading time appears to be spurious. It suggests that there is an increased reading time for the quantified sentence following refset reference. This would suppose that subjects returned to re-read the quantified sentence in this reference condition, which is counter-intuitive because, if anything, it is compset reference which is likely to be more difficult, as this involves reference to an implicit entity. Given this it seems reasonable to discount the observed effect as a chance event.

The failure to find any evidence of an effect invites the conclusion that the reference to an NP quantified by *only a few* is not processed as the sentence is read. While this is possibly true, because of difficulties associated with the interpretation of *only a few*, I inspected the off-line task completed by subjects immediately following eye-tracking, to be sure that the effects were not due to weaknesses in the experimental manipulation.

Results from the off-line post-test

The off-line results suggested that the experimental manipulation had failed for many of the materials. However, this was not due to the difficulty in reaching a compset-focused interpretation suggested by the pre-test, where less than 50% of continuations focused on the compset, but a failure to reach a refset-focused interpretation when this was predicted by the pragmatic manipulation. Overall, a mean 6.47 (sd=2.62) of a possible 12 refset-focused continuations were selected for materials in the refset-favouring context, and a mean 9.22 (sd=1.72) compset-focused continuations were selected for materials in the compset-favouring context. Results for one item were dropped from the analysis because of typographic errors in some of the experimental materials.

It was decided to re-analyse a reduced set of eye movement data for those items which showed the expected pattern of data in the off-line task. This was done by setting a criterion of 50% refset-focused continuations in the norm-matching condition, and 50% compset-focused continuations in the norm-violating condition. Twelve of the off-line materials matched passed both of these criteria, and the corresponding eye movement data was re-analysed.

The mean data across items for the off-line task is included in Appendix 4, and the items selected for re-analysis are marked by as asterisk.

Sentence reading time results for the reduced data set

The analysis of sentence reading times was repeated for the reduced data set, with separate analyses for the contextual, quantified and anaphoric sentence, using 2x2 within subjects ANOVAs.

This time the only effect was found on an analysis of the reading times for the anaphoric sentence, with a main effect of the pragmatic manipulation which was marginal by subjects ($F1(1,23)=3.64$, $p<0.067$; $F2(1,11)=1.39$, $p<0.264$). Although this failed to reach significance on the items analysis, this was most probably because of the reduced size of the data set, and the differences between means was in the same direction as those found in the subjects analysis. Reading times were longer when the quantified sentence violated norm expectations.

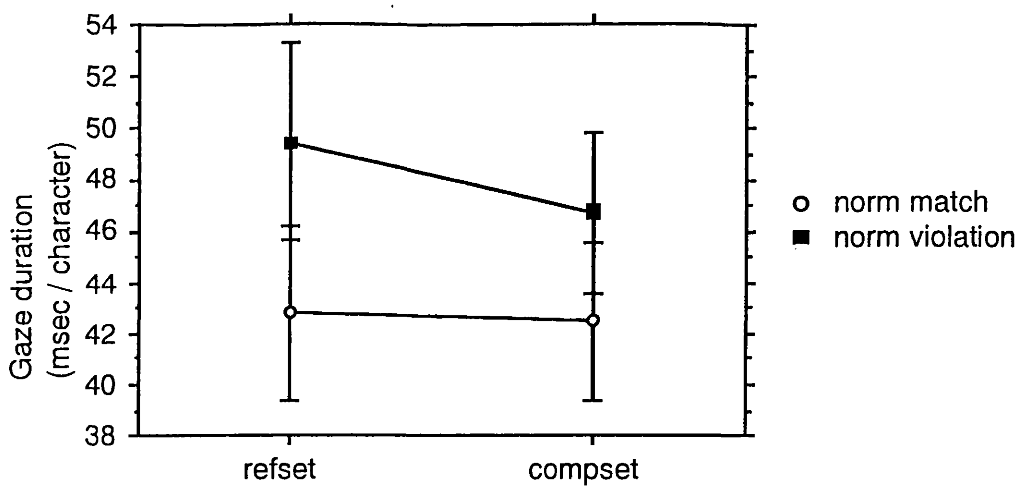


Figure 2: Mean reading time in msec / character for a sentence quantified by only a few in norm-violating and norm-matching conditions.

No other significant effects were observed, either for anaphoric sentence, or for analyses of the other sentence regions (all Fs<1.0).

Discussion

The analysis of the reduced data set still failed to find any evidence that anaphoric reference was contingent on the differential focus of refset and compset NP partitions following in norm-matching or violating conditions. What it did find was an overall increase in reading time for the anaphoric sentence when the preceding quantified sentence described a situation which violated norm expectations. If this proves to be a reliable effect, then it suggests that anaphoric reference was generally more difficult in this condition.

General discussion

The off-line production and on-line comprehension experiments investigated the same hypothesis: that focus is determined by both semantic and pragmatic constraints following quantification by *only a few*. All of the previous experiments have investigated the semantic constraints exerted on focus by monotone-increasing and monotone-decreasing quantifiers. However, formal observations, and data from a sentence-continuation task (Moxey and Sanford, 1987; 1993) suggests that *only a few* does not fit into either of these categories, nor does it exhibit a clear pattern of focus. Its interpretation appears to be more open to contextual constraint. My hypothesis was that *only a few* has the relatively simple semantic function of marking a set as small in size, but leaves it to pragmatics to determine the significance of this. It was predicted that when the marked small refset was consistent with pragmatic expectations, i.e. it was not unusually small given the described situation, then focus would be maintained on the refset. This meant that a subsequent pronoun would be interpreted as reference to this subset. In contrast, if the small refset violated expectations, i.e. it was unusually small, then the compset would be placed in focus, and there would be at least a tendency to interpret a subsequent pronoun as reference to this subset.

A production experiment in which subjects provided sentence completions that referred to the focused subset of a sentence quantified by *only a few* provided evidence in support of these claims. A significantly greater number of continuation sentences described the compset when the quantified sentence described a situation that violated expectations. Given this finding, it was concluded that set focus could be pragmatically influenced following quantification by *only a few*. The experiment also manipulated the form of connective between the quantified sentence and the anaphoric completion, and a significantly greater number of continuation for the connective *because*, and no effect for either *and, but* or a no connective condition. This finding suggested that since causal connectives like *because* mark the need to explain the described situation, they will also support compset focus. This appeared to be independent of the pragmatic manipulation.

The second experiment measured subjects' eye movements as they read sentences which referred to either the refset or compset of a sentence quantified by *only a few*. These quantified sentences were again manipulated to either match or violate expectations about the described situation. That is, when the described set (the refset) was not unusually small relative to expectations, then it would be maintained in focus, but when it was unusually small, then the compset would be placed in focus. It was predicted that reference to the unfocused·subset would appear anomalous and be more difficult to process. If this was an immediate and short-lived difficulty then it should have been observed as an increased reading time for the anaphor as it was first read. Alternatively, if the difficulty was experienced later, and represented difficulties in integrating the anaphoric sentence with the previous discourse, then there should have been reflected in reading times for the entire anaphoric sentence.

However, there was no difference in either the first pass reading times for the anaphor, or the global sentence reading times, when reference was made to the unfocused subset. That is, there was no evidence of an increased reading time when the anaphor described a property of the compset and the small refset matched expectations, nor for reference to the refset when this was marked as small relative to expectations. The failure to observe any effects due to the experimental manipulation can be explained in several ways.

First of all, it may be that reference is generally difficult to process when it is made to a sentence quantified by *only a few*. This consistent with the claim that focus is strongly determined by pragmatic constraints following this form of quantification. Chapter 3 reviewed a number of studies which demonstrated that inferences tend not to be processed during reading when they are dependent on pragmatics. For instance, McKoon and Ratcliff (1980) found that subjects do not infer that an actress has died when they read that she has fallen from the top of a 14th storey building. It was argued that these inferences are not made because they are defeasible, meaning that they can be ruled out subsequent text. For example, the actress may not have died when she fell off the roof because there was a safety net, or someone caught her at the last moment. Similarly, since focus appears to depend on pragmatics following quantification by *only a few*, and this is supported by sentence-completion results, it may be that it too is not processed during reading.

Alternatively, there may have been serious weaknesses with the experimental materials. One possibility is that the experimental materials may not have been sufficiently biased towards focus on the compset in the expectation-violating condition. This was identified as a problem during the pilot study. Subjects were asked to complete sentences which made pronominal reference to a preceding quantified sentence, and it was intended to select those materials for which there was a strong tendency to produce continuations that described the refset when the quantified sentence matched expectations, and continuations that described the compset when the quantified sentence violated expectations. However, it proved difficult to produce materials for which there was a strong tendency for subject to describe the compset in the norm-violating condition.

At the same time, there was no independent measure of the effectiveness of the referential manipulation used in the eye movement study. This meant that it was not clear whether the anaphor unambiguously described a property of the refset or compset in each of the referential conditions. There was some evidence of this from the post-test administered to subjects once they had completed the eye movement study. These subjects were presented with the experimental materials and asked to indicate which of the two alternative forms of anaphor were most appropriate in each of the pragmatic conditions. While there was 77% agreement that the anaphor which described a property of the compset was most appropriate when the quantified sentence violated expectations, there was only 54% agreement that the anaphor which described the refset was most appropriate when the quantified sentence matched expectations.

Such problems suggest that a general improvement of the experimental methodology is required before any conclusions are drawn about the processing of reference to sentence quantified by *only a few*. At the same time, the experiment can be further improved by stacking the odds in favour of compset focus in the norm-violating condition. One way of doing this would be to conjoin the quantified and anaphoric sentences by a causal conjunction like *because*, or include another compset-supporting operator such as a negative polarity item. The results from the sentence-completion study have already demonstrated that conjunction by *because* increases the likelihood that subjects will produce continuations that describe the compset. It may also increase the likelihood that the compset is placed in focus during the reading of the quantified sentence. This would again be consistent with the observation made in Chapter 3 that possible inferences are more likely to be made during reading when they have semantic support. There is other, more anecdotal reasons for including a negative polarity item. When designing the experimental materials, it was noticed that there was a tendency to include negative items in those sentences intended to produce compset focus. For example, (19) appears to produce a more focal compset than (20).

(19) Only a few of the weight-lifters managed to lift **even** the lightest dumbbell.

(20) Only a few of the weight-lifters managed to lift the lightest dumbbell.

Summary

There is off-line evidence that both semantics and pragmatics can constrain focus in some quantification conditions. The quantifier *only a few* appears to be used to mark an NP as a small set, and the significance of this is determined by pragmatics. When the smallness of the set is consistent with norm expectations focus is maintained on the refset partition of the quantified NP. However, when it violates these expectations, the compset is placed in focus. There is also evidence that other function words, like connectives, will interact with quantification to constrain focus. In particular, causal connectives like *because* increase the likelihood that subjects will produce continuations that describe the compset. This is because causal connectives mark the need to explain the situation described by the quantified sentence (Moxey and Sanford, 1987).

These effects were not replicated in an experiment which measured eye movements during reading. This can be interpreted as evidence that focus is not processed during reading when it is determined by pragmatics. However, there was also some evidence that weaknesses in the design of the experiment and experimental materials may have contributed towards the null results. Some alternative experimental designs were suggested.

Chapter nine:

Constraints on quantifier scope ambiguity

Introduction

This chapter makes a further digression to consider the interpretation of sentences containing more than one quantifier. The introduction describes the formal account of these sentences, which claims that they are ambiguous between a set of alternative logical interpretations, and that the extent of this ambiguity is determined by the structure of the sentence.

The chapter goes on to consider some possible accounts of the processing of these sentences. In particular, there are a number of theorists (e.g. Lakoff, 1971, 1972; Chomsky, 1976; Ioup, 1978) who claim that the preferred interpretation of these sentences is determined by structural information, and other researchers (e.g. Fodor, 1982; Kurtzman and MacDonald, 1993) who consider this as a processing constraint. However, other researchers (Ioup, 1978; Fodor, 1982; Sanford, 1990; Kurtzman and MacDonald, 1993) argue that there are more important semantic and pragmatic constraints on the interpretation of multiply-quantified sentences. According to some of these accounts, all three constraints (i.e. structure, semantics and pragmatics) may interact to determine the preferred interpretation of these sentences, and this may best characterised as a process of multiple constraint-satisfaction (Ioup, 1978; Kurtzman and MacDonald, 1993). That is, the various constraints may compete before the processing system finally settles on a best-fit interpretation.

The claims for semantic and pragmatic constraints are of particular interest because it appears that quantificational semantics may be at least partially responsible for resolving the scope ambiguity. The next chapter presents some off-line evidence of semantic and pragmatic constraints on scope resolution, and considers how it might be possible to investigate these constraints during normal reading.

Quantifier scope ambiguity

It is commonly observed (eg Allwood, Andersson and Dahl, 1977) that sentences which contain two quantifiers are also two-ways ambiguous, because they can be assigned one of two logically distinct scope interpretations. For instance, a sentence like (1) is ambiguous between the interpretations paraphrased by (2) and (3).

(1) Every man loves some woman.
(2) Every man loves some or other woman.
(3) Every man loves a single and particular woman.

The first of these requires that *every man* takes scope over *some woman*[30], and is represented in predicate calculus by a formula like (4), in which the relative scope of the universal ($\forall$) and existential ($\exists$) quantifiers is indicated by their order of mention. Because quantifiers on the left take scope over those on the right, the formula states that *for every man there exists a woman such that man loves that woman.*

(4) $\forall m\ \exists w\ (L(m, w))$

This scope relation is reversed for the other possible interpretation (as paraphrased by 3), so that *some woman* takes scope over *every man*, and can be logically represented using a formula like (5). This time the formula states that *there exists some woman such that every man loves that woman.*

(5) $\exists w\ \forall m\ (L(m, w))$

Given that the possible interpretations differ only in the order of the quantifier prefixes, this implies that quantifier scope ambiguity has a factorial complexity. This means that a sentence with three quantifiers will have six possible quantifier scope readings, and one with five quantifiers will have 120 possible quantifier scope readings. If this ambiguity is experienced as part of linguistic performance, i.e. during language use, then the processing of multiply-quantified sentences will rapidly prove intractable as the number of quantifiers increase, or else there must be constraints on the availability of alternative interpretations.

This is intuitively plausible: the difficulty in comprehending multiply-quantified sentences does appear to follow a steep incline as more quantifiers are added; while many of the possible readings are unavailable. For example, Hobbs and Shieber (1987) assert that only five of the six possible readings are

---

[30] Scope describes a dependence relation between functions, where one function is applied within the context, or scope, of another. I will sometimes talk about one NP exhibiting dominant scope, taking scope over another, or of an NP taking wide or widest scope (cf Fodor, 1982, for a discussion of the appropriate terminologies).

available for (6); and that only 42 of a possible 120 readings are available for (7). Since two of its five readings prove to be isomorphic, (6) can be understood to mean that representatives from the same company either each saw a set of samples, or saw the same set together; or that representatives from different companies either together or separately saw a set of samples. However, it requires considerable conscious effort to determine all of these readings, and it is extremely difficult to imagine many of the possible interpretations of (7).

(6) Every representative of a company saw most samples.

(7) Some representatives of every department in most companies saw a few samples of each product.

There are two views about the role of structural information in determining the possible readings of a multiply-quantified sentence. On one view (eg Jackendoff, 1972; May, 1977), structure determines the extent of the ambiguity, meaning that there will be as many possible readings as there are permutations of the relative quantifier order. These possible readings are then selected between by semantics and pragmatics. On the other view (Lakoff, 1971, 1972; Chomsky, 1976), structural information is responsible for determining the preferred interpretation of a multiply-quantified sentence. Lakoff makes particularly a strong version of this claim, according to which multiply-quantified sentences are unambiguous, and the correct logical interpretation has the same order of quantifier terms as the surface sentence. However, Chomsky takes a less extreme position, according to which the surface order of quantifiers determines a preferred reading, but the other possible logical interpretations are available as alternative readings.

Lakoff's claims are formalised as two syntactic constraints. The first constraint asserts that scope ambiguity only holds over quantifiers which reside in the same clause, and that quantifiers from a superordinate clause will take scope over those in the subordinate one. This means that a sentence like (8) has an unambiguous reading where there is a single set of texts which every philosopher has read.

(8) There are some books which have been read by all philosophers.

The second constraint states that the scope relation of quantifiers within the same clause is also unambiguous, and determined by the surface linear order of the quantifiers, just as the linear order of quantifiers determines scope in predicate calculus formulae. This means that the leftmost quantifier takes scope over those on the right.

The claim that the surface order of quantifiers determines either the sole or preferred reading of a multiply-quantified sentence has been criticised by Ioup (1978). In particular, she presents counter-examples to Lakoff's claim that linear order disambiguates these sentences. She argues that the sentence in (9) is consistent with the linear order hypothesis, because the first mentioned quantified NP, *a foreign student*, has dominant scope. This reading can be paraphrased as the claim that a particular student was persuaded to enrol in a number of different courses.

(9) I persuaded a foreign student to enrol in every course.

However, Ioup argues that (10) is a counter-example, which has a preferred reading in which the second mentioned quantified NP, *every course*, has dominant scope. This time the reading can be paraphrased as the expectation that some or other foreign student will enrol in each of the courses.

(10) I expected a foreign student to enrol in every course.

Given that surface linear order does not determine the preferred in reading of multiply-quantified sentences, Ioup proposes an alternative account according to which it is the relative grammatical position of quantifiers, and not their linear order, which determines scope. She describes a hierarchy (illustrated in Figure 1[31]) of grammatical categories which tend to take widest scope. At the top of this hierarchy are those constituents which contain the sentential topic. Ioup argues that the sentence topic is most likely to take wide scope in a sentence, and she notes that in English the topic NP is also the first mentioned one. This accounts for the considerable explanatory power of the linear order hypothesis. However, it is not a universal constraint, and there are other grammatical categories which can take wide scope. Below the sentential topic are those NPs which are both the deep and surface subject of the sentence, followed by those which are only the deep or surface subject. Then comes the indirect and direct objects, with the indirect object more likely to take scope over the direct one.

[31] Ioup describes grammatical categories in terms of deep and surface structures which are found in standard and extended theories of generative grammar, but not in current theories (Chomsky, 1981).

topic
deep and surface subject
deep subject / surface subject
indirect object
direct object

Figure 1: Ioup's hierarchy of grammatical categories most likely to receive highest scope

The evidence for this classification comes from intuitive judgements of scope preferences across a number of languages. However, only some of Ioup's English examples will be presented. First of all consider the sentences in (11) to (14).

(11) Every girl took a chemistry course.
(12) A chemistry course was taken by every girl.
(13) Every chemistry course was taken by a girl.
(14) A girl took every chemistry course.

Ioup argues that in (11) the universal quantified NP, *every girl*, has wide scope because it is both the surface and deep subject of the sentence, and that it tends towards wide scope in the passive version of this sentence (12) because the NP is still the deep subject. The tendency is less strong in (13) because *every chemistry course* is just the surface subject, and in (14) it falls under the scope of *a girl* because it is neither the surface or deep subject.

Similarly, (15) and (16) demonstrate that indirect objects take scope over direct objects. Although the linear order of the quantifiers is unchanged between the two sentences, *every child* takes highest scope in (15), and *a child* takes highest scope in (16).

(15) I told every child a story.
(16) I told every story to a child.

Given these observations, it seems likely that there is a more complex relationship between sentence structure and quantifier scope ambiguity than proposed by either Lakoff or Chomsky. The factorial of the number of quantified NPs can determine the extent of the logical ambiguity, but the surface linear order of quantified NPs need not correspond to the preferred logical interpretation. Ioup proposes that this is replaced by a more complex set of structural constraints.

Ioup also claims that semantic and pragmatic factors will help to determine the preferred reading of a sentence containing quantifier scope ambiguity, and that these constraints can over-ride structural preferences. In particular, the quantifier *each* will override structural constraints which favour an interpretation in which the NP quantified by *each* is under the scope of another NP. For example, although (17) has a structurally preferred reading in which the NP quantified by *some* takes wide scope, and means that a particular girl took each and every chemistry course; (18) has a different preferred reading in which the NP quantified by *each* has wide scope, and means that some or other girl took each and every course.

(17) Some girl took every chemistry course.
(18) Some girl took each chemistry course.

Ioup suggests that quantifiers can be hierarchically ordered by their tendency towards wide scope, in the same way that grammatical categories appear to form a hierarchy (see Figure 2). She places the universal quantifiers *each* and *every* at the top, then ranks the other quantifiers in terms of the size of set which they denote. Quantifiers which denote large amounts have a stronger tendency towards wide scope than those which denote small amounts. However, the scheme is weakened by its omissions. In particular, Ioup does not consider how negative quantifiers like *not many* and *few* might fit into it.

each
every
all
most
many
several
some
a few

Figure 2: Ioup's hierarchy of quantifiers which are most likely to receive highest scope

As with the hierarchy of grammatical categories, the evidence for this classification comes from cross-linguistic judgements. As before, I will only present some of Ioup's English examples in sentences (19) to (24). She claims that the preferred interpretation of (19) is one where there was a limited number of handouts and each of the pedestrians received one of them. However, as the second mentioned quantifier gets larger, the contrary scope reading, that each of the pedestrians received more than one handout, becomes dominant. This reading is strongly preferred in (23) and (24).

(19) Joan gave a few handouts to some pedestrians.
(20) Joan gave a few handouts to several pedestrians.
(21) Joan gave a few handouts to many pedestrians.
(22) Joan gave a few handouts to all pedestrians.
(23) Joan gave a few handouts to every pedestrian.
(24) Joan gave a few handouts to each pedestrian.

Kurtzman and MacDonald (1993) explore the further possibility that the thematic roles filled by quantified NPs will influence scope interpretation. They argue, following Jackendoff (1972) and Grimshaw (1990) that there is a hierarchy of thematic roles according to which those NPs which fill an Agent role will tend to takes scope over those NPs which fill an Experiencer role. Agent roles are associated with action verbs like *climb* or *hit*, and are filled by the NP that performs the verb action. Experiencer roles, however, are associated with verbs of perception such as *see* or *hear*, and are filled by NPs which participate in these experiences. Finally, those NPs which are either the Agent or Experiencer will tend to take scope over NPs which fill the Theme role, where the Theme is the recipient of the action or perception.

In terms of selecting preferred interpretations, this hierarchy of thematic roles makes the same predictions as Ioup's thematic hierarchy when applied to simple active sentence. This means that the leftmost NP will take scope over the one on the right. However, the thematic hierarchy also predicts that the preferred reading of a passive sentence is one in which the rightmost quantifier takes scope over the one on the left. It also predicts that there will be a stronger bias towards these interpretations when an active rather than a perception verb is used. For example, in (25) the action verb *climbed* has an Agent as its first argument and a Theme as its second; while the structurally equivalent (26) contains a perception verb (*saw*) which has an Experiencer as its first argument, and a Theme as the second. The thematic hierarchy predicts that (25) will be more biased to a wide scope universal reading than (26), because Agents show a stronger tendency to dominant scope than do Experiencers.

(25) Every kid climbed a tree.

(26) Every kid saw a tree.

In terms of the processing of scope-ambiguous sentences, both Ioup, and Kurtzman and MacDonald, propose that there is an interaction between structural constraints, the semantics (e.g. the semantic properties of quantifiers and verbs) will determine the preferred scope interpretation, although Ioup does not specify a mechanism for this interaction. Kurtzman and MacDonald suggest that the processing of multiply-quantified sentences can be explained by the same constraint-satisfaction model which MacDonald and her colleagues (eg MacDonald, 1994; MacDonald, Pearlmutter and Seidenberg, 1994) have applied to the processing of garden path sentences. This requires that preferences are encoded at the lexical level, and these compete and combine during syntactic processing, to eventually settle on a stable interpretation. Such a scheme would propose that quantifiers are encoded in terms of their preference for dominant scope, and the argument structure of verbs will also favour particular scope interpretations. The differential strength of these constraints will determine the outcome.

An alternative account of the relationship between syntax, semantics and pragmatics is proposed by Fodor (1982). Unlike the other theories, this is deliberately pitched as a processing account, and depends on some assumptions which are very similar to claims made in Chapter 1 about the mental representation of language. The first of these assumptions is that language comprehension consists of the construction of mental model analogues of the described situation. These are similar in form to the 'mental models' proposed by Johnson-Laird (1983), and specify the number of actors and actions in a situation. It was argued in Chapter 1 that mental representations of this type are performance correlates of formal semantic models.

The manner in which Fodor's version of these models represent mixed existential and universal quantification is illustrated in Figures 3 and 4. Figure 3 illustrates the wide scope existential, or $\exists\forall$ reading of (27), where the universally quantified NP is represented by an arbitrary number of tokens, each of which is connected by a line to a single representation of the existentially quantified NP. This line represents the sentence predicate *loves*. In contrast, Figure 4 illustrates a wide scope universal interpretation (or $\forall\exists$ reading), where each of the men love a separately specified woman.

(27) Every man loves some woman.

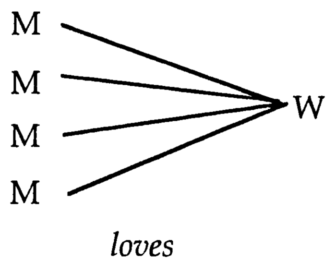


Figure 3: A model representation of a wide scope existential interpretation of *Every man loves some woman.*

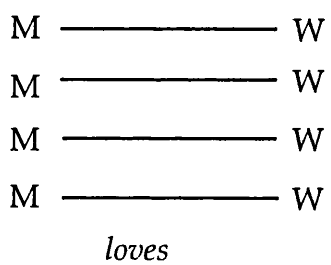


Figure 4: A model representation of a wide scope universal interpretation of *Every man loves some woman.*

Fodor assumes that syntax and semantics specify the manner in which these models are constructed. Quantifier semantics correspond to mental procedures which determine the form of representation for NP and predicate information, while syntax specifies the order and context in which these procedures are applied. For example, active and passive sentences require that NPs are introduced into models in a different order. So, if (27) was written in the passive voice, then it would be represented by a mirror image of Figures 3 or 4, with the tokens for *some women* on the left in the model. This is very similar to Montagovian claims about the isomorphism of syntax and semantics, which Johnson-Laird paraphrased as semantic ingredients mixed to a syntactic recipe.

According to Fodor, quantifiers divide into those which prompt a single representation of NP and predicate information, and others which induce multiple representations. It is the latter set which have the marked function, and only these which can be said to exhibit scope. By this account, scope is not a dependency between quantifiers, but the propensity of a quantifier to induce multiple representations. Existential quantifiers do not exhibit scope simply because they do not induce multiple representations; and other quantifiers can be distinguished by the range of their scope. Fodor suggests that this might explain why some quantifiers, like *each* and *every*, exert a stronger tendency towards wide scope than others (Ioup, 1978).

A quantifier like *all* will take scope over a sentence like (28) in one of three ways.

(28) All the children lifted a rock.

Either scope will be restricted to the NP it quantifies, with the interpretation that all of the children collectively lifted a specific rock; or that it will range across the first NP and the verb to mean that all of the children individually lifted the same rock; or finally, it will range across the entire sentence, and mean that all of the children lifted a different rock. Figures 5 to 7 illustrate the models which correspond to these three alternative readings.

In the first of these models (Figure 5), the lines converge prior to the verb, which indicates that scope is restricted to the quantified NP. In the second model (Figure 6), they converge after the verb, and there is no convergence in the final model because the quantifier has scope over the entire sentence.

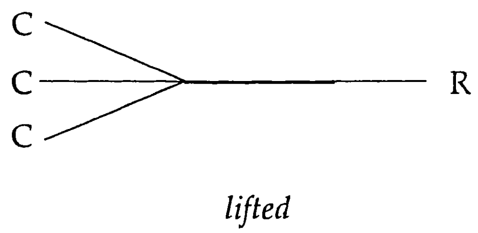


Figure 5: A mental representation for
all of the children collectively lifted the same rock.

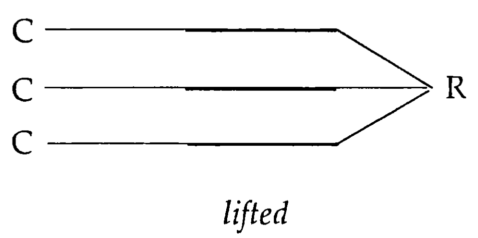


Figure 6: A mental representation for
all of the children lifted the same rock individually.

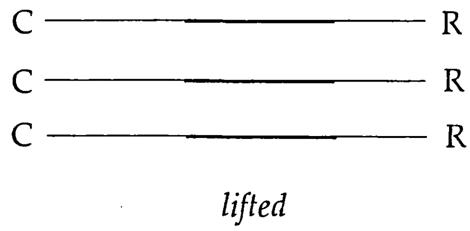


Figure 7: A mental representation for
all of the children lifted a different rock.

Fodor argues that universal quantifiers *each, every* and *all* differ in the extent to which they take scope over other sentence components, and it seems likely that the same is true of the other quantifiers in Ioup's semantic hierarchy. *Each* shows the strongest tendency towards a wide scope interpretation, which Fodor accounts for as the requirement that at least one constituent other than the quantified NP is multiply instantiated. So in terms of the models illustrated in Figures 5 to 7, a sentence like (29) must be interpreted as Figure 6, with a multiple instantiation of both the quantified NP and verb, if not Figure 7, where all of the constituents are multiply instantiated. This requirement is less stringent following quantification by *every,* and does not apply to *all.*

(29) Each child lifted a rock.

The semantic constraints exerted by multiple-inducing quantifiers interact with syntactic constraints on the order in which sentential constituents are entered into a mental model; which may also account for the preferred readings of multiply-quantified sentences. Fodor describes how a sentence like (30) is incrementally interpreted to produce a model representation.

(30) A child saw every squirrel.

The unmarked existential quantifier *a* is first of all interpreted as an instruction to represent a single child performing a single act of seeing. But this interpretation must be revised on encountering the multiple-inducing quantifier *every.* The least costly revision is one which produces multiple acts of seeing a squirrel, which corresponding to the preferred $\exists\forall$ reading of the sentence, as illustrated in Figure 8. This requires back-tracking to allow the universal quantifier to take scope over and induce multiple representations of the verb. The other possible interpretation, the ∀∃ reading, is more costly because it requires further back-tracking to produce multiple representations of both the verb and the first NP.

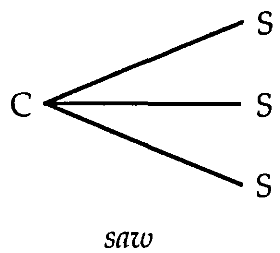


Figure 8: A mental representation of the ∃∀ reading of *a child saw a squirrel.*

Fodor's account of the preferred ∀∃ reading for a sentence like (31) is less satisfactory however.

(31) Every child saw a squirrel.

She argues that *every* causes the multiple representation of *child* and *saw*, and also *a squirrel*, because the indefinite article is unmarked. This produces a representation which is consistent with the ∀∃ reading of the sentence. However, Fodor also argues that it is difficult to reach the ∃∀ reading because the representational system uses a mental shorthand when producing multiple representations. Instead of instantiating an arbitrary number of NPs which are individually connected to other NPs, it represents one linked pair, with the mental footnote than the others should be represented in the same way. This shorthand representation, she argues, is more difficult to undo than the fully-instantiated model.

This is an ad hoc addition to her theory, and quite unnecessary. The difficulty in reaching the ∃∀ reading is better explained by the marked status of multiple-inducing quantifiers. It may simply be more costly to converge the multiple paths after then verb, than to multiple instantiate the second NP. Moreover, just how costly it is will depend on the type of quantifier used. For instance, the quantifier *each* will be less likely to allow convergence than *all.*

There is also evidence that quantifier semantics will interact with pragmatic constraints. For instance, (32) has an ∃∀ reading as its preferred interpretation, despite the multiple-inducing property of *all*. The ∀∃ reading is dispreferred because it is implausible that a soldier could surround a fort on his own.

(32) All soldiers surrounded a fort.

The substitution of *every* for *all* produces a synonymous sentence (see 33), and takes the plausible ∃∀ reading. However, the substitution of *each* into the sentence is unacceptable (see 34), because the strong multiple-inducing function of this quantifier over-rides pragmatic constraints to produce the implausible ∀∃ reading.

(33) Every soldier surrounded a fort.
(34) Each soldier surrounded a fort.

Sanford and Moxey (1995) consider some other instances where pragmatics can override semantic constraints. The preferred interpretation of (35) is one where each student is supervised by a particular tutor, although they may share their tutor with other students. In contrast, the identically structured (36) is understood to mean that each room has its own bath, and none are shared. In fact, it would cause consternation for a guest to find that it actually meant that every room on the floor shared a bath at the end of the corridor.

(35) Each student has a tutor.
(36) Each room has a bath.

The evidence of strong pragmatic constraints has consequences for theories which depend on off-line measures or intuitions about quantifier use. It is not clear for instance, where the line is drawn between semantic and pragmatic contributions to scope resolution in the sentence frame used by Ioup to justify her hierarchy of semantic constraints. For instance, the preferred interpretation of (37), where *all* takes wide scope, may depend on the greater plausibility of this interpretation compared to one where every pedestrian receives a single handout, which in total number *a few*. A proper study of semantic constraints on scope resolution must consider the possible contribution of pragmatics, and preferably tease these apart on-line.

(37) Joan gave a few handouts to all pedestrians.

Generalised quantifier theory

Most of the above observations of semantic constraints on scope resolution also appear in some recent formal semantic treatments of the interpretation of sentences with plural, co-ordinated or quantified NPs. The most important of these is Link's theory of the logic of plurals and mass terms (Link, 1983, 1987; also Lønning, 1987). Link argues that plural NPs have two possible representational states: one where they denote a collection of individual elements, and another where the NP is treated as an unindividuated whole. This is the same claim as made in Chapter 2, where it was observed that mass terms like *water* or *sand*, and group terms like *committee* or *army* are restricted to the latter type of representation, and argued that this is also the default representation for count nouns.

Link describes a sum operation which takes individual elements and conjoins them to produce an unindividuated structure. This procedure can be called during co-ordination, or is marked as a property of some predicates. In (38), for example, the use of *and* licenses a treatment of *John and Mary* as a single entity, which is supported by the predicate *met in Munich.*

(38) John and Mary met in Munich.

This can only be interpreted as predication over the conjoined pair, and not as a predicate which applies to the two individuals. Otherwise it would be possible to say that *John met in Munich and Mary met in Munich*, which is clearly unacceptable. Link describes this as collective predication, and there are a number of verbs which exhibit this property, including *surrounded, gathered in, dispersed, evacuated* and *scattered.* Collective verbs such as these cannot be used in sentences where the Agent NP is quantified by quantifiers like *each.* It has already been shown (in 34) that such a combination of *each* and *surrounded* appears anomalous.

Other predicates are ambiguous between collective and distributed readings. For instance, in (39) the predicate *lifted a stone* can either apply separately or conjointly to John and Harry. This ambiguity can be resolved by using a quantifier to mark one of the two readings (40).

(39) John and Harry *lifted a stone.*

(40) John and Harry *each* lifted a stone.

Lønning (1987) argues that quantifiers can also be categorised into those which do and do not permit collective readings. Some, including *every* and *each* allow only the distributive readings. Others like *two* or *an odd number of* are usually understood collectively. However, most quantifiers tend to take a collective reading, but can be understood distributively when this is warranted. This latter group includes *many, most, few, at least three* and *at most four.* Link suggests that we start out with an analysis in terms of sums, then should we happen to meet a distributive predicate we move down to level of individuated predication.

The generalised quantifier approach has similarities with the representational account proposed in Chapter 2. In both cases, it is claimed that the default mental representation of plural or quantified count nouns is an unindividuated mass, but that they can be represented as a set of individual members when necessary. Given that distributive quantifiers like *each* are used to mark this more individuated representation, now Link suggests that VP predicates are similarly constraining.

Given this similarity, it may also be possible to extend the representational account outlined in Chapter 2 to include co-ordination structures, which Link identifies as a device for mapping individuals into a collective representation. This appears to be a complement of the use of quantifiers to produce more fine-grained representations. According to this type of account, a quantifier like *a few* is used to map an NP into two subsets, which fill distinct roles in the underlying discourse, and place one of these subsets in focus. In (41), this means that the set of MPs is divided into those who attended the meeting, and some others who were absent. However, only the refset partition, those MPs who attended the meeting, or the superset, is available to pronominal reference.

(41) *A few* MPs attended a meeting.

A co-ordination structure like *and*, in contrast, maps individuals into a single discourse role, and licenses singular pronominal reference which will take the perspective of one of the individuals, and plural pronominal reference which takes the perspective of the conjoined pair. For example, continuations to (42) can refer to either Joan, James, or the conjoint pair.

(42) Joan *and* James went for a walk.

Link suggests that the form of predication will also determine the likelihood a co-ordinated interpretation: those like *gathered* force the conjoint reading, while other predicates like lifted a stone are ambiguous between the co-ordinated and uncoordinated interpretations. There is other evidence that the form of co-ordination plays a role. Sanford and Lockhart (1990; see also Sanford and Moxey, 1991) found that *and* was a better determinant of co-ordination than *with*. They also found that the co-ordinated individuals must have the same level of description. When one character was introduced by role description, and the other a named character, there was a lower tendency for plural reference than when both characters were introduced by either role description or name.

Summary

This section has concentrated on some theoretical accounts of quantifier scope ambiguity, and some attempts to produce processing accounts that are based on these observations. According to traditional formal accounts, the ambiguity arises from the possible linear order of the quantified terms when the sentence is translated into logical form.

Some theorists (e.g. Lakoff, 1971, 1972; Chomsky, 1976) have argued that multiply-quantified sentences are either disambiguated in favour of a particular logical interpretation, or have a dominant or preferred logical interpretation. In both cases the preferred logical interpretation has the same order of quantifiers as the surface form of the sentence. However, Ioup (1978) provides some counter-examples to these accounts, and proposes that they are replaced with a more complex structural constraint, according to which there is a hierarchy of grammatical categories which tend towards widest scope. This means that NPs in those categories at the top of the hierarchy will take scope over NPs that are in categories that are lower in the hierarchy.

Ioup also argues that quantifiers and pragmatics are important determinants of the preferred interpretation of a multiply-quantified sentence. Although she does not consider the pragmatic constraints, she does propose a hierarchy of quantifiers which tend to take wide scope. The quantifiers *each* and *every* are at the top of this hierarchy. Other researchers (e.g. Fodor, 1982; Link, 1987; Lønning, 1987) also argue that NPs quantified by each and every tend to take widest scope, and define this tendency in terms of either semantic properties (i.e. these quantifiers are distributive) or processing principles (Fodor, 1982). These researchers also argue that properties associated with particular verbs will determine the preferred interpretation of a sentence. Link and Lønning argue that some verbs are semantically marked as 'collective', which means that they can only be used to predicate a plural (or quantified) NP as a whole, and cannot be interpreted as a predicate of the individual members of that NP. According to this account, when a collective verb appears in an active doubly-quantified sentence, the preferred interpretation is one in which the subject NP takes scope over the object NP. Fodor makes the same predictions for sentences containing these verbs, but argues that the preferences arise from the plausibility of the alternative interpretations.

Of these accounts, only Fodor's is intended as a processing theory, and this has some important differences from the formal accounts. Fodor argues that doubly-quantified sentences are not processed in terms of alternative logical interpretations, but in terms of more fine-grained representations of the relationships between participants in the described situation. These representations are similar in similar in form to the Mental Models described by Johnson-Laird. Fodor argues that, in terms of processing, scope describes the propensity of a quantifier to produce a multiple instantiation of sentence constituents. For instance, the quantifier *each* requires that there is a multiple-instantiation of the NP it quantifies and at least one other sentence constituent. These quantificational constraints will interact with other factors, including plausibility, and structural constraints. According to Fodor, sentence structure is important to the extent that it determines the entry of entities into the mental representation.

In summary, the formal literature has identified structural and semantic determinants of the interpretation of scope-ambiguous sentences, and allow for the influence of pragmatics. Moreover, these have all been proposed as processing constraints, and Fodor (1982) has described a processing account in which each of these constraints have an important role. Others (e.g. Kurtzman and MacDonald, 1993) propose a less specific model in which the preferred interpretation of a multiply-quantified sentence is reached by a processing of multiple constraint satisfaction. That is, the various constraints compete until the processing system settles on a final interpretation. However, these accounts are based on a limited amount of experimental investigation. The following section describes those experiments which have been conducted, while the next chapter reports an experiment on the interaction of context and quantifier semantics to determine the final interpretation of a doubly-quantified sentence.

Experiments on quantifier scope ambiguity

There are few empirical studies of the interpretation of scope ambiguous sentences, but what studies there are have mostly addressed the role of structural information. One of the earliest of these was conducted by Johnson-Laird (1969), and is often cited as evidence for that the linear order of quantifiers in the surface form of the sentence will determine the preferred logical interpretation.

Johnson-Laird's subjects read a number of doubly-quantified sentences which were logically quantified by *some, every* or *no*[32], and included predicate negation in some conditions; and written in either the active or passive voice. Subjects had then to indicate which of a series of diagrams were consistent with the meaning of these sentences. Two example sentences are given in (43) to (45). The sentence in (43) illustrates one of the active sentence conditions where the first NP was quantified by *every,* and the second NP is quantified by *some.* In other conditions the first NP was quantified by *some* and the second quantified by *every,* or both NPs were quantified by either *every* or *some.*

(43) Every man knows some woman.

Finally, some conditions included negation. In these conditions, the quantifier *any* was paired with *some* and the VP was negated (e.g. 44). These quantifiers appeared in both possible orders.

(43) Any man does not know some woman.

Subjects also read passive versions of these sentences (e.g. 45) which again varied the order of quantifiers.

(45) Some woman is known by every man.

Johnson-Laird's main finding was that subjects selected those diagrams which matched the preferred reading, as predicted by the linear order hypothesis, and that active and passive sentences received different interpretations. At the same time, a low frequency selection of the possible but dispreferred reading was evidence against that the linear order of quantifiers did not fully disambiguate the sentences, but that it was possible to reach the dispreferred reading. However, the interpretation of these findings as support for the linear

[32] The quantifier *any* was also used to express NP negation, ie *Any woman is not known by some man.*

order hypothesis is tempered by Ioup's observation that linear order in English is confounded by a tendency of the subject NP to express the sentence topic. Also, since Johnson-Laird only considered the interpretation of logical quantifiers, it is not clear whether these findings transfer to non-logical or numerical quantification.

Gil (1982) used the same methodology to show that group readings are available for multiple numerically quantified sentences. His subjects read a number of sentences like (46), which is repeated below, then indicated which of a set of diagrams corresponded to the meaning of the sentence. He found that subjects selected those diagrams which represented both the two scope differentiated interpretations, and the two group ones.

(46) Two examiners marked six scripts.

While both of these studies concur with some of the theoretical claims made in the previous section, the off-line nature of the task precludes any strong conclusions about the processing of these sentences. For instance, Johnson-Laird observes that subjects found the task difficult (Johnson-Laird, 1969, pp 10), while the time which his subjects took to classify a sentence indicates that considerable conscious effort was required (the mean classification time was 52.4 seconds). Normal sentence comprehension takes place within a much shorter interval, and without the same allocation of processing resources; which means that it is possible that the observed effects will not occur on-line. Similarly, the group readings observed by Gil may be reached via knowledge about plausible scenarios for the sentence, and calculated when mapping the sentences onto diagrams, and not during sentence processing.

Both Gillen (1992), and Kurtzman and MacDonald (1993) used measures of sentence reading time, or the time taken to judge the acceptability of a sentence, to investigate the on-line processing of doubly-quantified sentences. Gillen presented subjects with sentences like (47) and (48), where the direct and indirect objects were quantified and produced a scope ambiguity between a quantified and indefinite NP. The quantifier was one of *some, several, many, most, all, every* or *each*. This sentence was then followed by a continuation sentence (49) which made either singular to plural reference to the indefinite NP. Subjects read these materials sentence by sentence, and at their own pace.

(47) Susan gave a recipe to QUANTIFIER friend(s).
(48) Susan gave QUANTIFIER friends a recipe.
(49) The [recipe was / recipes were] for Hungarian goulash.

Gillen found an overall increase in reading time for the quantified sentence when the quantifier was in the indirect object position (48), and a difference between the quantifier types which she attributed to a faster reading time following quantification by *all*. She also found a faster reading time for continuation sentence which made singular reference to the indefinite NP. However, these results have limited significance. The finding that quantifier first sentences are read quicker than quantifier second ones was predicted by Fodor. With the latter case the processor must backtrack to multiply-instantiate the indefinite NP. Yet this is contradicted by the finding of easier singular reference to the indefinite NP. That suggests that it is not multiply-instantiated.

There is an alternative explanation for the faster singular reference. An inspection of Gillen's materials shows that most of them are ambiguous between type and token readings. For example, (47) can either mean that Susan gave a separate recipe to each of her friends, or gave each of them a copy of the same recipe. This latter interpretation is consistent with a multiple instantiation of the indefinite NP but also with singular reference. Another example of this confound is given in (50) and (51). In (50), James is probably handing out a set of tickets which are almost identical to each other and for the same event. It is then perfectly felicitous to refer to these tickets by definite NP (eg (51).

(50) James handed a ticket to some shoppers.
(51) The ticket was for a charity.

It is even more likely that differences in the reading time for singular and plural reference was due to the fact that there are more characters in the plural reference sentence. Gillen did not carry out any transforms (e.g. a ms/character transform) to control for the difference in sentence length. This may also account for the differences in reading time between the quantified sentences.

Kurtzman and MacDonald used similar materials to Gillen, but asked their subjects to indicate whether the referential sentence was a reasonable discourse continuation for the quantified sentence. This followed the same rationale as the Gillen study: plural definite NP reference is felicitous when the antecedent quantified NP is within the scope of a universal quantifier. Kurtzman and MacDonald used quantified sentences which were disambiguated by the inclusion of the post-determiners *same* and *different* (but see Carlson, 1987, for a treatment of these operators) as a baseline.

Kurtzman and MacDonald conducted three experiments: one to investigate the processing of simple active multiply-quantified sentence, another to investigate the corresponding passives, and a final study where the scope ambiguity resided within a prepositional phrase. All three studies used the universal quantifier *every* and the indefinite article *a*. The first two experiments also manipulated the type of verb used: either action verbs like *climbed* or perception verbs like *seen*.

In the first of these studies subjects read sentences like (52) and (55), followed by a sentence which made either singular or plural reference to the indefinite NP. Kurtzman and MacDonald found a general preference for the interpretation where the leftmost quantifier has wide-scope. This meant that (53) and (56) were the favoured continuations. However, this result was modulated by stronger evidence of this preference when *a* was the leftmost quantifier; and the dispreferred interpretation was more available for perception verbs. Kurtzman and MacDonald suggest that the last of these effects was due to a tendency for agents rather than experiencers to take scope over themes. They argue that this accords with Ioup's hierarchy of grammatical constituents.

(52) Every kid climbed [a | the same | a different] tree.
(53) The trees were full of apples.
(54) The tree was full of apples.

(55) [A | The same | A different] kid climbed every tree.
(56) The kid was full of energy.
(57) The kids were full of energy.

These results support suggest that the surface linear order of quantifiers can determine the preferred interpretation of a doubly-quantified sentence, but do not disambiguate the sentence since the low frequency acceptability of the dispreferred reading suggests that it was still available. However, the second experiment, which examined passive versions of the sentences, did not support these findings. This time there was no evidence of any preferred reading, and the two continuation sentences were judged equally acceptable. Moreover, because the frequency of acceptable judgements was not much lower than that for the active sentences, the results cannot simply reflect a greater difficulty in interpreting the passive form.

The third experiment examined sentences like (58) and (61), followed by sentences which made singular or plural reference to the indefinite NP. The control conditions were equivalent sentences which used *each* and *the* in place of *every* and *a*. This took advantage of strong lexical semantic preferences for a wide scope interpretation of these operators (see next section for a discussion of *each*).

(58) George has [every | each] photograph of [an | the] admiral.
(59) The admirals were quite famous.
(60) The admiral was quite famous.

(61) George has [a | the] photograph of [every | each] admiral.
(62) The photographs were quite famous.
(63) The photograph was quite famous.

As with the passive manipulation, the results failed to support the claim that the surface linear order of quantifiers determines the preferred reading, but instead suggested a clear preference for the rightmost NP to take wide scope. This meant that (60) was the preferred continuation of (58) and (63) the preferred continuation of (61), whatever the quantifier type or order. Even the strong lexical preferences of *each* and *the* were swamped by this preference.

Kurtzman and MacDonald conclude that there are two structural principles at work. One of these favours leftmost wide scope. This might be a linear order constraint, or due to the relative topicality of constituents. The other principle favours rightmost wide scope and may be explained by a thematic hierarchy. The first of these principles dominates when processing simple actives, but it is the latter which is dominant during the processing of passives. However, Kurtzman and MacDonald suggest that the two principles are balanced during the processing of passives, which explains the absence of any preferred reading. There may also be lexical semantic constraints exerted by quantifiers and verbs - and the form which these might take is considered in the next section.

The experimental studies , it seems, are either inconclusive about the processing of doubly quantified sentences, or suggest that there are several competing principles at work. The Gillen results, however, are flawed by a type/token ambiguities in the quantified sentences, and while the Kurtzman and MacDonald materials were less prone to this flaw, they are limited in their manipulation of quantification types. Finally, neither Gillen, nor Kurtzman and MacDonald control for the pragmatic influences noted by Fodor (1982) and formal semanticists like Link (1983, 1987) and Lønning (1987).

Chapter ten:

Experiment Eight:

Semantic and pragmatic constraints on the interpretation of quantifier scope ambiguity.

Introduction

The most obvious conclusion to the previous chapter is that very little is known about the processing of quantifier scope ambiguities. There is a hotchpotch of theories and principles, some of which are easily falsifiable, and others which make testable predictions. Kurtzman and MacDonald (1993), for instance, find some evidence of competing syntactic and semantic principles: where scope dominance depends on the linear order of quantifiers, and also on the thematic roles filled by the quantified NPs. This is particularly interesting because it is consistent with recent theories of parsing which also point to an interaction between structural processing principles and argument structure (eg Britt, 1994).

There is also the suggestion that some quantifiers are marked for scope dominance (Ioup, 1978; Fodor, 1982). This seems particularly true of the quantifier *each*; while Ioup describes a hierarchy of quantifiers which tend to take a dominant scope reading. However, at least one on-line study of the processing of quantified sentences (Gillen, 1992) has unsuccessfully manipulated this variable. This failure may indicate that quantifier semantics do not impact on scope resolution during on-line processing, or it might reflect flaws in Gillen's experimental design. For instance, I noted that many of the Gillen materials were ambiguous between type and token readings of one of the quantified NPs, which will have reduced the likelihood of observing any experimental effects.

The following experiment will test Ioup's hierarchy of quantifier semantics, but use an off-line sentence-continuation task. This will provide a more sound footing for subsequent on-line studies. If the off-line task does verify Ioup's hierarchy, then it leads to the question of whether the same effects can be observed in a time-limited task, and further questions about the loci of quantifier function. It also avoids the risk of failing to find any effects in a replication of the Gillen study, but still not knowing what happens off-line.

The present study also includes a pragmatic manipulation since there is evidence of pragmatic constraints on scope resolution, and a risk that these have contaminated Ioup's judgements on quantifier function. I observed that one of the possible scope interpretations was more plausible that the other in some of Ioup's quantificational conditions; and this may have overridden quantifier preferences (eg Sanford and Moxey, in press). The confound can be avoided by controlling for plausibility, or by manipulating pragmatics. The present study used a bland experimental material (1) to avoid problems with plausibility, and embedded these sentences in contexts which favoured one of the two scope interpretations.

(1) QUANTIFIER reporter(s) interviewed a survivor.

Method

The experiment used two off-line tasks: a sentence-continuation task, and a second task where subjects had to produce a line-drawing of the relationship between the two quantified NPs. Although the second of these tasks is more difficult, it is easier to categorise the scope interpretation from a drawing than a sentence-continuation.

Design

Two between subjects variables were manipulated. The experimental passages contained a sentence (1) in which the subject NP was quantified by one of six quantifiers: *each, every, all, most, many, several* or *some.* This sentence was embedded into one of two contexts, one which favoured a wide-scope interpretation of the quantified NP, and another which favoured a wide-scope interpretation of the indefinite NP. Subjects saw only one item in one of the quantification and pragmatic conditions.

Only a single type of passage was used, which means that it was not possible to generalise findings across texts. However, the results provided a first approximation about the use of context and quantification in the interpretation of scope ambiguity, and will motivate further experiments to test whether it is possible to generalise from these findings.

Subjects

The study used 312 unpaid subjects who were students at the University of Glasgow.

Materials

One quantified sentence (1) was used for the experiment, and embedded in a passage which favoured one of the scope interpretations. Passage (2) was intended to favour an interpretation where a number of reporters interviewed the same survivor. This corresponds to a collective, or wide-scope indefinite NP interpretation.

(2) The CNN news-team were the last on the scene of the aircrash, and the authorities had already organised a press conference. The CNN reporters joined the other journalists, and positioned their cameras in front of the podium. **QUANTIFIER reporter(s) interviewed a survivor.**

The other passage (3) favours an interpretation of the quantified sentence where a number of reporters separately interviewed their own survivor. This is a distributed, or wide scope quantified NP reading.

(3) The CNN news-team were the first on the scene of the aircrash, and determined to find the best eye-witness before any other news-teams arrived. The CNN reporters split up and worked their way through the wreckage. **QUANTIFIER reporter(s) interviewed a survivor.**

The form of quantification was manipulated within each of these pragmatic conditions by substituting one of *each, every, all, many, some* or *a few* for QUANTIFIER.

Procedure

Subjects were given two sheets of paper and instructed to complete the task described on the first page before turning to the second. They were given the sentence-continuation task first, and instructed to "Please read the short passage given below, then add one or two sentences to continue the story." They then read one of the experimental passages with the scope-ambiguous sentence in one of its quantifier conditions.

On a second page subjects were asked to draw the relationship described in the doubly-quantified sentence. They were instructed to "Please read again the passage on the previous page, and draw the diagram that you fell best represents the relationship between the participants of the final printed sentence (not your own final sentence)." Three diagrams of numerically-quantified sentences were provided as examples. The first (Figure 1) described a collective mapping and the second (Figure 2) a distributed mapping. The third (Figure 3) diagram described a mixed mapping.

Two pensioners got on a bus.

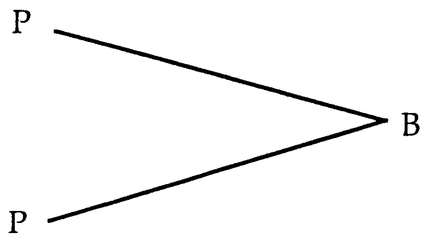


Figure 1: Example of a collectively-mapped diagram.

Four angler caught a fish.

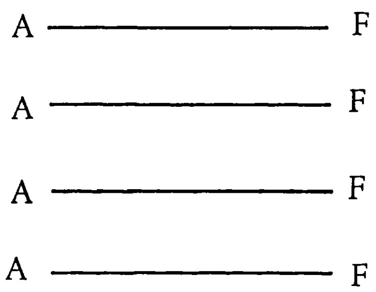


Figure 2: Example of a distributed-mapped diagram.

Three students passed two exams.

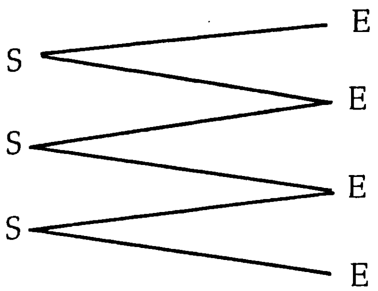


Figure 3: Example of a mixed-mapped diagram.

Results

The results for both the Diagram and Sentence-continuation task were categorised in terms of the mapping they described, either a **collective** mapping with a single instantiation of the indefinite NP, a **distributed** representation with multiple instantiations of the definite NP with a one-to-one mapping between reporters and survivors, or a **mixed** representation of a more complex situation.

Diagram Results

For a diagram to be classed as collective there had to be a single instantiation of the indefinite NP and a multiple instantiation of the quantified NP, with each of the quantified NP tokens attached to the indefinite NP token. This also included diagrams with additional and unattached quantified or indefinite NP tokens.

Distributed diagrams had to have a multiple instantiation of both NPs, and a one-to-one relationship between them. Diagrams of this form, but with additional unattached tokens were also classed as distributed. The mixed category comprised those diagrams that were neither collective or distributed, but described some complex intermediate position. A fourth category of **unclassed** diagrams was also added. This covered those diagrams where subjects had not understood (or followed) the instructions and produced diagrams which could not be classed as either collective, distributed or mixed. The mean results for these categories across context and quantifier conditions is summarised in Table 1.

This categorised data was analysed using an analysis of variance by chi-square (Winer, 1991). The collective and distributed categories were analysed separately because of their dependence; but the results for the mixed category could not be analysed as the expected cell frequency fell below the minimum of 5.

The analysis of collectively-classed diagrams identified significant main effects of context type ($\chi^2(1) = 6.44$, $p<0.05$), with a greater incidence of collective diagrams in the collective contextual condition, and of quantifier type ($\chi^2(5)=60.95$, $p<0.001$), but no evidence of an interaction.

A residual analysis (Siegel and Castellan, 1988) was used to determine whether the frequency of collectively-classed diagrams differed from chance for each of the quantifiers. This was significantly less collective diagrams for the quantifiers *each* (Residual = -3.91, $p < 0.05$) and *every* (Residual = -3.37, $p < 0.05$), and significantly more collective diagrams for the quantifiers *some* (Residual = 3.49, $p < 0.05$) and *a few* (Residual = 3.49, $p < 0.05$). However, the frequency of collectively-classed diagrams did not differ from chance for either *all* (Residual = -1.20, $p > 0.05$) or *many* (Residual = 1.50, $p > 0.05$).

Context	Quantifier	Collective	Distributed	Mixed	Other
Collective	Each	8	9	9	0
	Every	9	8	8	1
	All	17	3	4	1
	Many	25	0	0	1
	Some	25	0	1	0
	A few	25	0	0	1
Distributed	Each	1	16	9	0
	Every	3	13	9	1
	All	7	10	4	4[33]
	Many	14	8	3	0
	Some	25	0	0	1
	A few	25	1	0	0

Table 1: Scope categories for diagrams.

The parallel analysis of diagrams in the distributed category produced a significant main effect of context type ($\chi^2(1) = 11.53, p<0.001$), with more distributed diagrams in the distributed context condition, and a significant main effect of quantifier type ($\chi^2(5)=46.71, p<0.001$). Again there was no evidence of an interaction. An analysis of residuals (Siegel and Castellan, 1988) was used to determine if the frequency of distributively-classed diagrams differed from chance. There were significantly more distributive diagrams for the quantifiers *each* (Residual = 4.06, $p < 0.05$) and *every* (Residual = 2.87, $p < 0.05$), but significantly less distributive diagrams for the quantifiers *some* (Residual = -3.37, $p < 0.05$) and *a few* (Residual = 3.07, $p < 0.05$). The frequency of distributive diagrams did not differ from chance for the quantifier *all* (Residual = 0.50, $p > 0.05$) or *many* (0.99, $p > 0.05$).

Continuation Results

The continuation results were categorised by the number of survivors referred to by subjects. A collectively mapped interpretation requires that there is a single representation of the second NP. This will be reflected in continuations that refer to "he", "she", "the only survivor", or "the survivor". Continuations

[33] All noted the possible ambiguity.

that did so were classed as singular references. The following text fragments (4 to 6) are examples of singular reference continuations.

(4) The survivor said that the crash was one of the most harrowing experiences he had had to deal with. The shock had not yet sunk in that his wife was dead.

(5) They were interested as to what he could remember about the time just before the crash.

(6) The man was too shocked to give more information.

In contrast a distributed representation requires that more than one survivor was interviewed, and continuations which referred to a plural number of survivors were categorised as multiple survivor references. Continuations that described the reporters as returning to compile separate interview data were included as multiple references. Again, (7) and (8) are examples of such continuations.

(7) Each of the survivors said it as horrendous . . .

(8) The CNN reporter was last to his feet to ask one of the female survivors the question that was on everyone's lips.

Continuations that referred to survivors other than those interviewed, or described the actions of the reporters without either directly or indirectly describing the number of survivors were not categorised. The categorised continuations are summarised in Table 2.

Context	Quantifier	Single Survivor	Multiple Survivor	Unclassed
Collective	Each	4	19	3
	Every	8	14	4
	All	16	6	5
	Many	24	2	0
	Some	20	0	6
	A few	21	0	5
Distributed	Each	2	23	1
	Every	3	19	4
	All	11	10	5
	Many	11	12	3
	Some	17	0	9
	A few	21	1	4

Table 2: Scope categories for continuations.

These category results were analysed using a Chi-Square ANOVA. The analysis of single survivor references produced significant main effects for context ($\chi^2(1)=4.96$, $p<0.05$), and quantifier $\chi^2(5)=41.14$, $p<0.001$). There was no evidence of an interaction. An analysis of residuals (Seigel and Castellan, 1988) was used to determine if the frequency of single survivors continuations differed significantly from chance for each of the quantifiers. There were significantly fewer of these continuations for the quantifier *each* (Residual = -3.96, $p < 0.05$), and *every* (Residual = -2.98, $p < 0.05$), but there were significantly more of these continuations for the quantifiers *some* (Residual = 2.08, $p < 0.05$) or *a few* (Residual = 3.06, $p < 0.05$). However, the frequency of single-survivor continuations did not differ from chance for the quantifiers *all* (Residual = -0.13, $p < 0.05$) or *many* (Residual = -0.06, $p > 0.05$).

An analysis of the multiple survivors category also produced main effects for context ($\chi^2(1)=5.53$, $p<0.05$) and quantifier ($\chi^2(4)=81.13$, $p<0.001$), and no interaction effect. An analysis of residuals (Seigel and Castellan, 1988) was again used to determine if the frequency of multiple survivors continuations differed significantly from chance for each of the quantifiers. There were significant more multiple-survivor continuations for the quantifiers *each* (Residual = 5.78, $p < 0.05$) and *every* (Residual = 3.64, $p < 0.05$), but there were significantly fewer continuations for the quantifiers *some* (Residual = -4.21, $p < 0.05$) or *many* (Residual = -3.97, $p < 0.05$). The frequency of multiple-survivor continuations did not differ from chance for the quantifiers *all* (Residual = -0.40, $p > 0.05$) or *many* (Residual = -0.88, $p > 0.05$).

Discussion

The main experimental findings supported the hypothesis that both quantifier semantics and discourse context can contribute to the resolution of quantifier scope. An analysis of both the diagrams and continuations produced by subjects suggested that the scope interpretation of the quantified sentence was determined by the form of quantification and the context of the sentence. There was a strong tendency to interpret a sentence according to which the subject NP was quantified by either *each* or *every* as distributive, in which the subject NP took scope over the object NP, and there were multiple instances of the object NP. This resulted in diagrams where there was a one-to-one relationship between instances of the subject NP and instances of the object NP. Similarly, there was a greater tendency to produce continuations which referred to many or plural instances of the object NP (which was always indefinite).

This is consistent with the widespread claim that *each* and *every* are inherently distributive (e.g. Vendler, 1967; Ioup, 1978; Fodor, 1982; Link, 1987). This appears to be a very strong constraint. Moreover, those few continuation sentences which were classed as collective in these conditions, often described a temporal distribution. That is, although the quantified sentence in (21) was interpreted to mean that each of the reporters interviewed the same survivor, they each interview the survivor at different times. Such an interpretation is consistent with Fodor's claim that the quantifier *each* must take scope over the quantified NP and at least one other sentence constituent. In this case, the quantifier takes scope over the verb *interview* to produce an interpretation in which there are many acts of interviewing the same survivor.

(21) Each of the reporters interviewed a survivor.

In contrast, there was a strong tendency to interpret a sentence in which the subject NP was quantified by either *a few* or *some* as collective, according to which the object NP took scope over the subject NP, and there was only a single instance of this object NP. This resulted in diagrams where there was a many-to-one relationship between instances of the subject NP and a single instance of the object NP, and continuation sentences which referred to a single instance of the indefinite object NP. The quantifiers *all* and *many* did not exhibit a preference for either scope interpretation.

At the same time it was found that a manipulation of the discourse context could bias the interpretation of the scope-ambiguous sentence. When the discourse context was designed to favour the collective interpretation, there was a significant likelihood that subjects would produce diagrams in which there was a many-to-one relationship between instances of the subject NP and a single instance of the object NP; and a significant likelihood that continuations would describe single instances of the object NP. In contrast, when the discourse context was designed to favour the distributed interpretation, there was a greater tendency for subjects to produce diagrams with a one-to-one mapping between the subject and object NPs, and to produce diagrams which referred to multiple instances of the object NP.

The results across the quantificational conditions can be interpreted in line with Ioup's (1978) claim that quantifiers have a hierarchical tendency towards wide-scope. This could account for the strong tendency of *each* and *every* towards a wide-scope interpretation, and the contrasting weak tendency of *a few* and *some*. In fact, the almost exclusive narrow scope interpretation of sentences quantified by these can be explained by placing the indefinite article higher in the hierarchy (see Fodor, 1982). However, the problem with this explains is explaining where the hierarchy comes from.

Another possibility is that only some quantifiers, such as *each* and *every* are marked for wide scope (e.g. Link, 1987; Lønning, 1987). Scope interpretations of sentences containing the other quantifiers might be determined by syntactic and semantic factors like topic position and thematic roles which were not investigated in this experiment, but shown to be an important factor by Kurtzman and MacDonald (1993).

Quantifier scope resolution during reading

As already noted, there is an extensive and often contradictory literature on quantifier scope ambiguity, little of which has been empirically tested. Not that this means that the resolution of quantifier scope ambiguity is an uninteresting question - because it raises a number of issues which are central to psycholinguistics. For instance, the claim that the surface linear order of quantifiers determines the preferred interpretation of a multiply-quantified sentence is a ruled-based account which is in a similar vein to the minimal attachment and late closure accounts of parsing. The fact that it does not appear to be wholly supported by the existing empirical data (Gillen, 1992; Kurtzman and MacDonald, 1993) has parallels with current research into parsing where there is a debate between rule-based accounts (e.g. Rayner and Frazier, 1982) and a constraint-satisfaction approach (e.g.) MacDonald, Pearlmutter and Seidenberg, 1994), in which constraints until before the processing system settles on a particular structural decision.

Both Kurtzman and MacDonald, and Ioup (1978) suggest that scope resolution might also be a constraint satisfaction problem, where syntax, semantics and pragmatics compete to determine the scope interpretation. However, they are vague about the relative contributions of these factors, and divided over the nature of the semantic constraints. Ioup's suggestion of a hierarchy of quantifiers which tend towards widest scope is only partly supported by present experimental findings, while Kurtzman and MacDonald suggest a similar hierarchy of thematic roles. This means that active sentences containing verbs like *climbed*, where the subject NP is a actor, will be more inclined towards a collective or wide scope universal reading than when a verb like *saw* is used, where the subject NP is an experiencer.

Other verb constraints may be more pragmatic in origin. Fodor (1982), for instance, observes that some verbs like *surrounded* are marked for a collective reading, while other VP frames like *broke a leg* or *sipped some wine* strongly favour a distributed interpretation where each member of the subject NP separately performs the described act.

The interaction of these constraints during reading could be studied using the self-paced reading methodology employed by Gillen, and Kurtzman and MacDonald. Subjects would be presented with a discourse fragment like (22) which contains a scope ambiguous sentence where one of the NPs is indefinite, and a second sentence which refers to this indefinite NP by either singular or plural definite NP. Singular reference will be felicitous when the quantified sentence has a narrow scope interpretation, and there is a single instantiation of the second NP. Plural reference is felicitous when the sentence has a wide scope interpretation and there are multiple instantiations of the second NP. A manipulation of quantifier and syntactic structures will produce a clearer picture of how these interact to constrain the interpretation of scope-ambiguous sentences.

(22) *QUANTIFIER* of the kids climbed a tree.
The **[tree was | trees were]** full of apples.

Paterson and Edden (in preparation) use this self-paced reading methodology to examine the interaction of quantifier and verb constraints on the interpretation of doubly-quantified sentences. In one experiment (23) they investigate singular and plural definite NP reference to sentences where there is a scope ambiguity between an indefinite NP and one quantified by *all*. The sentences also include a verb which is marked for a collective reading (eg. *shared*) or unmarked (eg *carried*). According to Link (1987) and Lønning (1987), the quantifier *all* is unmarked, so singular and plural reference should be equally felicitous. However, because *shared* is marked for a collective interpretation, it should favour singular reference. It should be noted, however, that there are other plausible readings, including one in which each of the drug addicts shared the syringe with at least one other person[34]. If subjects reached this interpretation then either singular or plural reference would be plausible, which predicts that there would be no difference in reading time for the two continuation sentences.

(23) *All* of the drug addicts **[shared | carried]** a dirty syringe.

---

[34] This reading was suggested by Alan Garnham.

[The syringe was | syringes] were used to inject heroin.

The study will also follow Kurtzman and MacDonald (1993) in using materials like (24) and (25) as control conditions, where (24) is the baseline for a collective interpretation, and (25) is the baseline for a distributed interpretation.

(24) *All* of the drug addicts shared the same dirty syringe.
The syringe was used to inject heroin.

(25) *All* of the drug addicts carried a different dirty syringe.
The syringes were used to inject heroin.

A second experiment (26) will substitute the marked distributive quantifier *each* for *all*. This should favour plural reference when used alongside the unmarked verb, but conflict with the collective preference of the marked verb.

(26) *Each* of the drug addicts **[shared | carried]** a dirty syringe.
[The syringe was | syringes] were used to inject heroin.

Other studies will examine the interaction of these constraints for passive versions of the doubly-quantified sentences.

Chapter eleven:

Summary and conclusions

Formal and psychological approaches to semantics

The experiments reported in this thesis have made a general exploration of the processing of natural language quantification during reading. The motivation for this was made clear in Chapter 1, where I argued for a parallel between formal and psychological approaches to natural language semantics. In particular, a correspondence was drawn between the model representations used by formalists and the model-like mental representations proposed by some psycholinguists. Others have made the same point, and drawn the same parallels, particularly Johnson-Laird (1982, 1983), whose Mental Models theory is in the spirit of formal semantics but with an awareness of the performance restrictions required by a psychological implementation. Mental Models are partial model representations, and communication - as both Johnson-Laird (1982) and Hall-Partee (1979) observe - is successful to the extent that the participants in a discourse share the same partial model. Quantifiers are one example of linguistic operators which control the structure of these mental representations, and so should be seen as a means of maintaining shared partial models.

At the same time it was noted that there are drawbacks to the Mental Models theory, particularly Johnson-Laird's assumption that the membership of a plural noun can be represented by an arbitrary number of tokens. This assumption is not appropriate for all types of NP. Mass and group nouns do not readily decompose into a set of member elements, and quantified NPs may not do so either. This is unfortunate, because Mental Models is a powerful theory which can explain human performance on a variety of reasoning tasks (cf. Steedman and Johnson-Laird, 1978; Johnson-Laird, 1983; Johnson-Laird and Byrne, 1991). My solution, which was detailed in Chapter 2, is to incorporate aspects of Mental Models theory within an account of discourse representation and processing proposed by Sanford and Garrod (1981; Garrod and Sanford, 1982; 1991).

The result is an account where there are two levels to the mental representation of text. There is an explicit focus representation of those entities which are mentioned in the text, and a mapping between explicit focus and an implicit focus representation of the discourse roles in the underlying scenario. It is assumed that plural nouns take an unindividuated representation in explicit focus, unless there are quantifier or pragmatic constraints to the contrary. This means that they are represented by a single token in explicit focus, and map onto a single discourse role in the underlying scenario. Mapping the complex explicit focus representation onto a simple underlying disourse role is intended to express the idea that the unindividuated members of a plural noun fill the same single role. For example, *the MPs* in (1), is represented by a single explicit focus token and fills a solitary implicit focus role-slot, as illustrated in Figure 1. The upshot of this account is that pronominal reference is felicitous when it refers to the plural noun as a collective whole, and anomalous when it is made to member elements.

(1) **The MPs** attended a meeting.

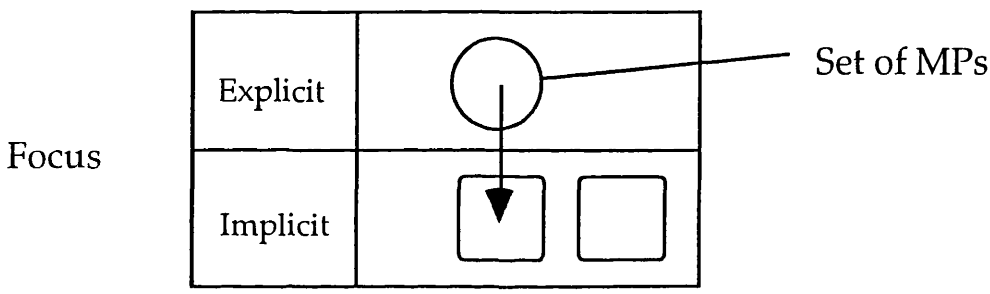


Figure 1: The mental representation of *the MPs attended a meeting*

The function of quantifiers

Some forms of quantification are used to produce more fine-grained representations, and to focus processing attention onto particular aspects of the discourse model. Quantification by either *a few* or *few*, for instance is used to divide the plural noun into refset and compset partitions, which are respectively true and false of the sentence predicate. These subsets have distinct explicit focus representations and map onto separate discourse roles in the underlying scenario. Mapping them into separate discourse roles marks the fact that the subsets are contrasted in terms of their discourse roles. However, only the refset appears to be mentally represented following quantification by *a few*, but both subsets are represented following quantification by *few*.

This suggests that while the mental representation of a sentence like (2) will be similar to that for the unquantified plural (e.g. 1), i.e. similar to Figure 1, the mental representation of the sentence in (3) will have two explicit focus tokens, one for the refset of MPs who attended the meeting, and another for the compset who were absent. These tokens are presumed to map onto separate role-slots in implicit focus, as illustrated in Figure 2.

(2) **A few of the MPs** attended a meeting.

(3) **Few of the MPs** attended the meeting.

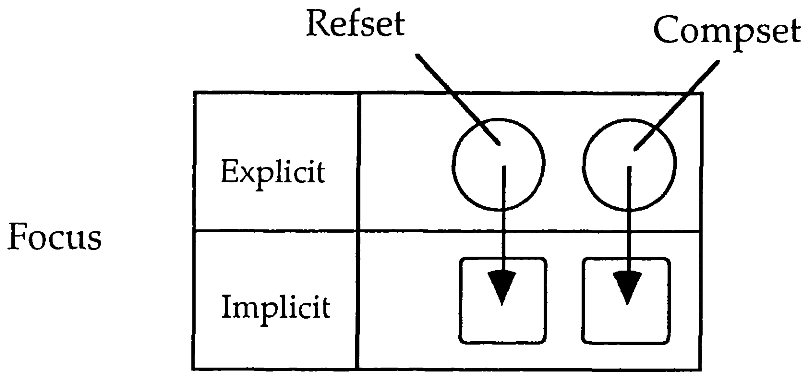


Figure 2: The mental representation of
few of the MPs attended a meeting

This account has consequences for the processing of pronominal reference. It suggests that pronominal reference to a sentence quantified by *a few* will be interpreted as reference to the refset partition. However, if the pronoun is intended as reference to the compset, then this will appear anomalous, since the compset is not available as an antecedent. To overcome this anomaly, it is necessary to add the compset into the mental representation and revise the initial reference assignment from the refset to the compset. It was predicted that this would be a costly process and be evident as an increased reading time for the anaphoric sentence. In contrast, since the mental representation of a sentence quantified by *few* has an explicit representation of both the refset and compset, this suggested that a pronoun could be plausibly interpreted as reference to either subset. However, off-line studies of language production (Moxey and Sanford, 1987) suggested that the compset is the preferred antecedent under these quantificational conditions, presumably because the compset is more focal. This suggested that during comprehension, there would also be a preference for interpreting the pronoun as reference to the compset. However, if the pronoun was intended as reference to the refset, then this would not appear anomalous - just dispreferred - and the assignment would be more easily revised.

The first two experiments, reported in Chapter 4, used a sentence-by-sentence self-paced reading task to test whether pronominal reference is contingent on quantifier focus during normal reading. The first experiment measured the reading time for sentences which made plural pronominal reference to either the refset or compset of a sentence quantified by either *a few* or *few*, while the second experiment replicated this experiment with the quantifiers *many* and not *many*. It was predicted that there would be greater difficulty in processing, and therefore a longer reading time for the anaphoric sentence when it referred to a property of the unfocused subset. That is, there would be a longer reading time for anaphoric sentences which referred to the compset of a monotone-decreasing quantified sentence (i.e. one quantified by either *a few* or *many*), and to the refset of a monotone-decreasing quantified sentence (i.e. one quantified by either *few* or *not many*).

The experimental results supported the claim that monotone-increasing and -decreasing quantifiers can be distinguished in terms of focusing properties, but did not support the claim that monotone-decreasing quantifiers focus on the compset. There was a longer reading time for the anaphoric sentence when it made reference to the compset following quantification by *a few* or *many*, and a longer reading time for reference to the compset following quantification by *not many*. However, the difference in reading time for reference to either the refset or compset was not reliable following quantification by *few*. This suggested that both the refset and compset were plausible antecedents of pronominal reference following monotone-decreasing quantification, although there was a preference to interpret the pronoun as reference to the compset following quantification by *not many*.

The results were interpreted as evidence that quantifier focus occurs during and that pronoun resolution is contingent on the form of quantification. It was more difficult to integrate an anaphoric sentence with prior discourse when the sentence described a property of an unfocused subset. However, it was not possible to conclude that the increased reading times were due to the detection and revision of an initial and anomalous attachment to the focal subset, or simply indicative of greater processing difficulty for reference to the unfocused antecedent. The first of these explanations suggests that there is a point in the processing of the anaphoric sentence when anomalous reference it detected. Furthermore, it was predicted that referential anomaly detection would occur at the earliest possible point, when the anaphor refers to a property of the unfocused subset.

The locus and time course of referential anomaly detection

Three studies of eye movements during reading were conducted in an attempt to identify the locus of referential anomaly detection. These experiments were conducted within the context of a number of studies which have demonstrated the on-line processing of pronominal reference - meaning that reference is processed as the pronoun is read - and particularly a study by Garrod, Freudenthal and Boyle (1994) which found evidence of the immediate impact of prior discourse focus and pragmatic inference. Garrod et al found an immediate increase in First Pass reading time when the role-description of the pronoun was incongruent with the discourse role of the focused antecedent. It was predicted that the current experiment should also find evidence of immediate anomaly detection when the anaphor described a property of the unfocused subset.

The first two studies replicated the previous self-paced reading experiments, and found a longer Total reading times for the anaphoric sentence when the anaphor described a property of the compset following quantification by either *a few* or *many*. Similarly, there was a longer Total reading time for the anaphoric sentence when it described a property of the refset following quantification by *few*, but no difference in Total reading time for sentences which referred to a property of the refset or compset following quantification by *not many*. However, there was evidence of a more localised increase in Total reading time when the anaphor referred to the refset (*their presence* in 4) than when it referred to the compset (*their absence*). This was interpreted as evidence that reference to the refset of a monotone-decreasing quantified sentence produces only a short-lived difficulty, and that this is occurs because the refset is a necessary part of the mental representation of the sentence.

(4) /*Quantifier* of the MPs went to a meeting./
/**Their [presence | absence]** /helped /the meeting run more smoothly.

There was no reliable evidence of referential anomaly detection during the initial processing of the anaphoric sentence. According to a measure of First pass reading time, i.e. the fixation time on the first visit to a text region, there was no significant difference in reading time for either the anaphor or subsequent regions for text across quantifier and reference conditions. However, there was some evidence of difference in the frequency of First pass regressions from particular regions of text. There were more regressions overall from the anaphoric region (*their presence/absence* in 4) following quantification by *few*, which may indicate general difficulties in processing reference under this type of quantification. At the same time, there was an increased number of regressions from the subsequent verb region (*helped* in 4) following anomalous reference to an NP quantified by *a few*, and from the final region (*the meeting run more smoothly*) for anomalous reference to an NP quantified by *many*. However, when the data from the two experiments were combined, there was no evidence of a common pattern of results within the monotone-increasing and -decreasing conditions, which suggested that the results were either spurious, or dependent on individual properties of the quantifiers. This meant that the measure of First pass regressions did not provide evidence of processing disruption that was due to the focusing properties of the quantifiers.

Chapter 6 attempted to account for the failure to observe punctuate referential anomaly detection in terms of the differences between the present experimental design and that employed by Garrod et al (1994). In particular, they sited their anomaly on an intransitive verb and investigated singular pronoun reference, while the present design sited the anomaly on a noun and used plural pronouns. The first of these differences may have contributed to the defeasibility of the pragmatic inference which detects the referential anomaly. It was noted that the processor may not commit to an anomalous interpretation on encountering the anaphoric manipulation, because it is always possible that it will turn out to be acceptable at some point downstream in the text. For instance, what is an apparently anomalous reference to the unfocused compset in (5) turns out to be felicitous reference to the refset by the end of the sentence. This means that unless the manipulation itself is sufficiently strong, then there will be no evidence of a punctuate effect.

(5) A few of the MPs went to a meeting.
Their absence would have angered the Whips.

Using plural pronouns in place of singular pronouns may also have been a mistake because of ambiguities associated with plural pronoun reference. In the present materials the plural pronoun is always ambiguous between reference to either one of the subsets, or reference to the superset. There may be a delay in the processing of pragmatic inference until this referential ambiguity has been resolved.

The third eye-movement study, reported in Chapter 7, incorporated some of the proposed revisions. The referential manipulation was moved to the quantified sentence, the referential anomaly was cited on an intransitive NP (*gambled recklessly* in 6), and a causal conjunction *so* was used to conjoin the quantified and referential sentences. These revisions had the added advantage of allowing a comparison of identical regions (*they gambled recklessly*) across experimental conditions.

(6) / *Quantifier* of the men were **[careful | careless]** with their winnings, / so / **they** ***gambled*** / ***recklessly*** / until / the money / was gone.

There was evidence of difficulty in processing reference to the compset of a sentence quantified by *a few*, with longer reading time for both the quantified and anaphoric sentences under theses conditions. There was also evidence of a general difficulty in processing reference following quantification by *few*, with further evidence that refset reference produced a more long-term processing difficulty under these conditions. The Total reading times for the anaphoric sentence and early regions of the anaphoric sentence were generally long following quantification by *few*, and often as long as the Total reading times for reference to the compset following quantification by *a few*. However, the Total reading times for the later regions of the anaphoric sentence were significantly longer when reference was made to the compset following quantification by *few*. This was interpreted as evidence in favour of an account in which only the refset is a plausible antecedent following monotone-increasing quantification, but while both the refset and compset are plausible antecedents following monotone-decreasing quantification, the compset is the favoured one.

Despite the efforts to localise the referential effect, there was still no evidence of punctuate referential anomaly detection according to either First pass measures of reading time or regressions. There was evidence of an overall longer reading time for the anaphoric region (the pronoun + verb + adverb) when reference was made to the compset. This suggested that reference was initially processed independently of focus information, and certainly independently of focus information following quantification by *few*. There was also evidence of an increase in First pass reading time for the NP and VP regions at the end of the anaphoric sentence when reference was made to the compset following quantification by *a few*. This suggested the referential anomaly had a delayed impact on processing, however it occurred so late in the sentence as to be indistinguishable from second pass (i.e. Total reading time) effects.

Given that all three eye-tracking experiments failed to find any evidence of immediate referential anomaly detection, it can be reasonably concluded that the difficulty in processing reference to the unfocused subset of a quantified sentence is not a localised phenomena, but a more global difficulty in integrating the anaphoric sentence with prior discourse. However, it could still be argued that the experimental manipulations were insufficiently powerful to find these effects, and some further experiments outlined in in Chapter 6 may provide a definitive answer. The first of these proposed experiments will use stronger predictors of compset focus to enhance the experimental contrast: *hardly any* will be substituted for *few* (and contrasted with the monotone-increasing *some*), while *and so* will be substituted for the connective *so*. Off-line data suggests that both of these operators produce a large incidence of compset focus (Sanford et al, in press). An example material is illustrated in (7).

(7) / **[Some | Hardly any]** of the men were **[careful | careless]** with their winnings, / and so / **they** ***gambled*** / ***recklessly*** / until / the money / was gone.

Another study will avoid ambiguities associated with plural pronoun reference by comparing singular pronominal reference to NPs quantified by the monotone-increasing *one* and the complex exception-marking expression *all but one*. Again, an example material is illustrated in (8).

(8) / **[One | All but one]** of the men was **[careful | careless]** with his winnings. / **He** ***gambled*** / ***recklessly*** / until / the money / was gone.

Finally, a third study will directly test the alternative theories by comparing the reading time of sentences where apparently anomalous reference is made felicitous by placing the adversative connective *nevertheless* either at the beginning or the end of the anaphoric sentence. An example material is given in (9).

(9) / **[A few | Few]** of the men were **careful** with their winnings. / [Nevertheless /] they *gambled* / *recklessly* / until / the money / was gone [/ nevertheless].

Delayed referential processing predicts no difference between the reading time of sentences with *nevertheless* in these two sentence positions. However, an immediate initiation of referential processing predicts a temporary anomaly when nevertheless is in the sentence final position. This should result in a longer reading time compared to the case when nevertheless is placed at the beginning of the sentence. Replicating the experiment with monotone-increasing and monotone-decreasing quantification should also determine whether these quantificational conditions alter the time course of referential processing.

Pragmatic constraints on quantifier focus

The first set of experiments considered the function of monotone-increasing and monotone-decreasing quantifiers during comprehension. However, it was noted that there was a third category of quantifier, which are described as non-monotonic because they do not license directional scalar inferences. One such quantifier is *only a few*, which is a complex quantifier composed of the negative and monotone-decreasing *only* and the positive and monotone-increasing *a few.* Moxey and Sanford (1987; 1993a) found that this has a default for refset focus, but can result in compset focus when supported by conjunction semantics, or other pragmatic constraints. In particular, Moxey and Sanford (1987) report an increased number of compset focused sentence continuations when the quantified sentence described a situation which is norm-violating.

Chapter 8 reports two experimental investigations of this claim. The first experiment was a large scale sentence-continuation task where subjects produced continuations to a quantified sentence which either matched (eg 10) or violated (eg 11) norm expectations. These sentences were either followed by a plural pronoun, or conjoined to a plural noun using *and, but* or *because.* There was a greater tendency for compset-focused continuations in the norm-violating condition (11), and when this was supported by the causal conjunction *because*.

(10) *Only a few* of the weight-lifters managed to lift **the heaviest** dumbbell.
(11) *Only a few* of the weight-lifters managed to lift **the lightest** dumbbell.

The second experiment transferred this pragmatic manipulation to a reading task, and measured subjects' eye movements as they read the same quantified sentences, but embedded in a short passages like (12).

(12) At the gym.
Some expert weight-lifters from the gym held a weight-lifting competition. Only a few of the weight-lifters managed to lift the [**heaviest / lightest**] dumbbell. Their [**strength / weakness**] was a surprise to their friends.

This experiment added the referential manipulation used in the previous studies, where an anaphoric NP (*Their [strength / weakness]*) is used to refer to a property of either the refset or compset partition of the quantified NP. When the quantified sentence describes a norm-matching situation, then the refset should be in discourse focus, and the preferred antecedent of pronominal reference. This predicts a global increase in reading time when the anaphor describes a property of the compset, and perhaps also evidence of more punctuate referential anomaly detection. However, when the quantified sentence describes a norm-violating situation, then the compset should be more focal, and reference to the compset should be at least as easy as refset reference. However, no significant differences were found on an initial analysis of the eye movement data.

The data set was then reduced on the basis of an off-line forced choice task which had been administered immediately following the eye-movement task, and although the re-analysis still failed to find any significant differences, there was some evidence of an overall increase in Total reading time for the anaphoric sentence in the norm-violating conditions. If this is a replicable finding then it suggests that reference resolution is more time-consuming in this condition because of the complex pragmatic inferences required both to instantiate the compset, and to compute a match between the anaphor and antecedent.

It may be that the failure to find any convincing reading time effects is due to the weakness of the experimental manipulations. After all, the off-line task only produced an about 50% incidence of compset-focused continuations in the norm-violating condition. A more effective test of the on-line hypothesis would include a causal conjunction like *because* to increase the likelihood of compset focus.

Semantic and pragmatic constraints on quantifier scope ambiguity
The final section of the thesis broadened the treatment of natural language quantification to include quantifier scope ambiguity. This ambiguity arises when two or more quantifiers (including the definite and indefinite articles)

appear in the same clause. The standard account claims that a multiply-quantified sentence is ambiguous between logical representations which differ only in the order of the quantifier prefixes. For example, sentence (13) is ambiguous between the predicate calculus formulae in (14) and (15). The first of these can be paraphrased to mean that *for each boy there exists a girl who is loved by that boy.* The second means that *there exists a particular girl who is loved by every boy.*

(13) Every boy loves some girl.

(14) ∀boy ∃girl Loves (boy, girl)

(15) ∃girl ∀boy Loves (boy, girl)

Chapter 9 reviewed some theoretical approaches to this problem, which variously argued for syntactic, semantic and pragmatic constraints solutions. What empirical studies have been conducted (Gillen, 1991; Kurtzman and MacDonald, 1993) find little evidence that syntax can resolve the ambiguity, but do identify semantic factors. Chapter 10 reported an off-line study which demonstrated an interaction of semantic and pragmatic constraints on the ultimate interpretation of a scope-ambiguous chapter, and argued for an similar approach to this question as taken in the earlier consideration of quantifiers and focus.

Following the approach taken by some formal semanticists (Link, 1983, 1987; Lønning, 1987) and also by Fodor (1982), it was suggested that some quantifiers, such as *each*, have a marked preference for a individuated representation of the quantified NP, and some verbs, such as *surrounded*, have a marked collective interpretation. Other quantifiers and verbs are ambiguous between collective and distributed interpretation. These constraints can also be modelled within my representational account.

First of all consider the quantifiers. Given a sentence like (16) where both the quantifier and predicate are unmarked, the quantified NP (*the soldiers*) is assigned an unindividuated representation in Explicit focus as the default, and maps into a single role slot in Implicit focus - as illustrated in Figure 3. The lack of individuation means that the sentence predicate (*visited a castle*) holds over the quantified NP as a collective whole: so there is one act of visiting, and a single instantiation of the indefinite NP (*a castle*). The sentence is interpreted to mean that a contextually specified group of soldiers collectively visited a castle.

(16) All of the soldiers visited a castle.

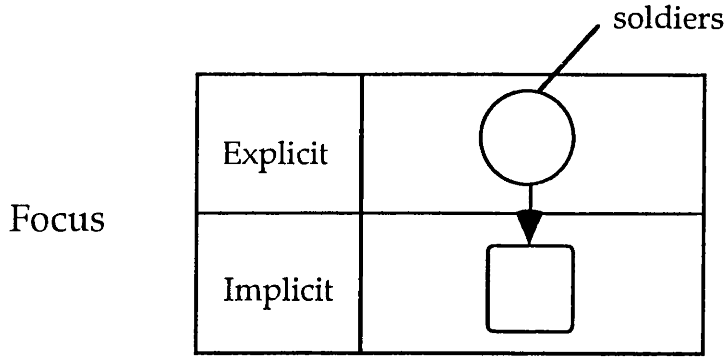


Figure 3: The mental representation of
all of the solders visited a castle.

However, when *all* is replaced by a marked distributive quantifier like *each* (17), then the quantified NP takes an individuated representation where a number of tokens which stand for individual soldiers are mapped into separate role-slots - as illustrated in Figure 4.

(17) Each of the solders visited a castle.

This individuation of the quantified NP means that the sentence predicate now applies to the individual set members, and the sentence is understood to either means that they separately visited the same castle, or that they visited different castles. Mapping the individuated entities into separate discourse roles emphasises the contrast between the characters. A sentence like (17) asserts that the soldiers visited a castle in a distnctive manner, i.e. by either visiting different castles or visiting at different times.

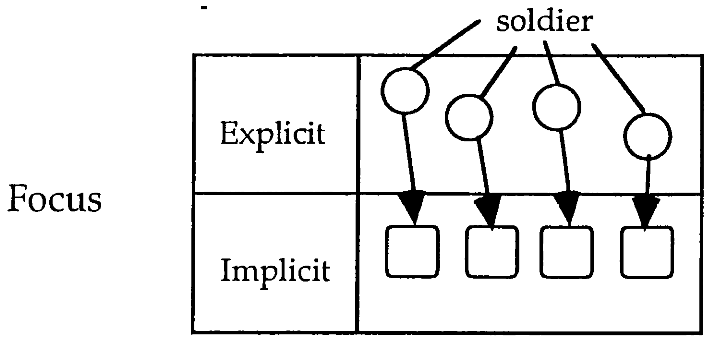


Figure 4: The mental representation of
each of the solders visited a castle.

Fodor (1982) suggests that these two readings are differentiated by the extent to which the quantifier takes scope over the rest of the sentence. In the first of the two readings the quantifier only takes scope over the verb, to produce a multiple instantiation of the acts of visiting. However, in the second reading it takes scope over the whole of the VP predicate to multiply instantiate both the acts of visiting, and the visited object.

This does not explain how the scope of the quantifier might be restricted. One possibility is that there is an interaction with verb constraints. When the verb is marked for a collective interpretation, it may block the individuating function of the quantifier. For example, the antagonistic function of the quantifier and verb in (18) appears to produce a preferred reading where soldiers perform multiple acts of gathering at a single castle.

(18) *Each* of the soldiers *gathered in* an ancient castle.

Of course this is all speculation, but it does produce testable claims. A first series of experiment (Paterson and Edden, in preparation) is intended to investigate the interaction of quantifier and verb constraints on the interpretation of multiply-quantified sentences. It will follow a similar methodology as Kurtzman and MacDonald (1993) and measure the acceptability of singular and plural definite NP reference to the indefinite NP of a scope-ambiguous sentence like (18). However, where they manipulated syntax and quantifier semantics, this study will manipulate quantifier and verb semantics.

Conclusion

Semantic processing has been neglected largely neglected in the sentence processing literature in favour of syntax. However, if the Formal semantic approaches are correct, then it may not be so easy to separate the two, and semantics may exert stronger constraints on initial sentence processing than has been supposed. The experiments reported in this thesis were intended as a starting point. They extend a set of observations made by Moxey and Sanford (1987, 1993a), and show that quantificational semantics can direct referential processes. A further experiment was reported to consider whether the resolution of quantifier scope ambiguity can be explained in a similar manner. Quantifier scope ambiguity is becoming a more central concern for the sentence processing community (cf. Frazier, 1995), and is a topic where syntax and semantics converge. Even broader approaches might consider whether the quantificational semantics considered in this thesis plays a role in other topics which in the past have been considered to be a wholly syntactic concern, such as the interpretation of garden-path sentences.

References

Allwood, J., Andersson, L-G., and Dahl, O. (1977). *Logic in Linguistics.* Cambridge: Cambridge University Press.

Anderson, A., Garrod, S., and Sanford, A. J. (1983). The accessibility of pronominal antecedents as a function of episodic shifts in narrative texts. *Quarterly Journal of Experimental Psychology, 35A*, 427-440.

Anscombre, J-C., and Ducrot, O. (1986). Argumentation et informativité. In Michel Meyer (Ed.), *De la Metaphysique à la rhetorique.* Brussels: Editions de l'Université de Bruxelles.

Bach, K. (1982). Semantic nonspecificity and mixed quantifiers. *Linguistics and Philosophy, 4*, 593-605.

Barclay, J.R. (1973) The role of comprehension in remembering sentences. *Cognitive Psychology, 4*, 229-254.

Barker, C. (1994). Group terms in English: Representing groups as atoms. *Journal of Semantics, 9*, 69-93.

Barwise, J., and Cooper, R. (1981). Generalised quantifiers and natural language. *Linguistics and Philosophy, 4*, 159-219.

Bosch, P. (1988). Representing and accessing focused referents. *Language and Cognitive Processes, 3*, 207-231.

Bower, G.H., Black, J.B., and Turner, T.J. (1979). Scripts in memory for text. *Cognitive Psychology, 11*, 177-220.

Bransford, J. D. and Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 717-26,

Bransford, J.D., Barclay, J.R. and Franks, J.J. (1972) The abstraction of linguistic ideas. *Cognitive Psychology, 3*, 331-350.

Britt, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language, 33,* 251-283.

Caramazza, A., Grober ,E. H., Garvey, C., and Yates, J. (1977). Comprehension of anaphoric pronouns. *Journal of Verbal Learning and Verbal Behaviour. 16,* 601-609.

Carlson, G. N. (1987). Same and different: some consequences for syntax and semantics. *Linguistics and Philosophy, 10,* 531-565.

Carpenter, P. A., and Just, M. A. (1972). Semantic control of eye movements in picture scanning during sentence-picture verification. *Perception and Psychophysics, 12,* 61-64.

Carpenter, P. A., and Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review, 82,* 45-73.

Carpenter, P. A., and Just, M. A. (1976). Models of sentence verification and linguistic comprehension. *Psychological Review, 83,* 318-322.

Carpenter, P. A., and Just M. A. (1977). Reading comprehension as eyes see it. In M. A. Just and P. A. Carpenter, *Cognitive processes in comprehension.* Erlbaum: Hillsdale, NJ.

Carreiras, M. (1993). Minimalist misconceptions of mental models. PSYCOLOQUY 3(64) reading-inference-.7

Carreiras, M., Garnham, A., and Oakhill, J. (1993). The use of superficial and meaning-based representations in interpreting pronouns: Evidence from Spanish. *European Journal of Cognitive Psychology, 5,* 93-116.

Chafe, W. L. (1972). Discourse structure and human knowledge. In J.B. Carroll, and R. O. Freedle (Eds.), *Language Comprehension and the aquisition of knowledge.* Washington DC: Winston.

Chafe, W. L. (1976). Giveness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic.* Academic Press: New York.

Champaud, C., and Bassano, D. (1987). Argumentative and informative functions of French Intensity modifiers 'presque' (almost), 'a peine' (just, barely) and 'a peu pres' (about): An experimental study of children and adults. *European Bulletin of Cognitive Psychology, 7*, 605-631.

Chase, W. G., and Clark, H.H. (1971). Semantics in the perception of verticality. *British Journal of Psychology*, 62, 311-326.

Chomsky, N. (1976). *Reflections on language.* Fontana: London.

Chomsky, N. (1980). *Rules and representations.* Basil Blackwell: Oxford.

Clark, H. H. (1976). *Semantics and comprehension.* The Hague: Mouton.

Clark, H. H., and Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3*, 472-517.

Clark, H. H., and Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics.* Harcourt Brace Jovanovich: New York.

Clark, H. H., and Sengul, C. J. (1979) In search of referents for noun-phrases and pronouns. *Memory and Cognition, 7*, 35-41.

Cloitre, M., and Bever, T. G. (1989). Linguistic anaphors, levels of representation, and discourse. *Language and Cognitive Processes, 3*, 293-322.

Cohen, J. D., MacWhinney, B., Flatt, M., and Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments & Computers, 25*, 257-271.

Corbett, A. T., and Chang, F. R. (1983). Pronoun disambiguation: Accessing potential antecedents. *Memory and Cognition, 11*, 283-294.

Cotter, C. A. (1984). Inferring indirect objects in sentences: Some implications for the semantics of verbs. *Language and Speech, 27*, 24-45.

Dell, G.S., McKoon, G., and Ratcliff, R. (1983). The activation of antecedent information during the processing of anaphoric reference in reading. *Journal of Verbal Learning and Verbal Behavior, 22*, 121-132.

Dooling, D.J. and Lachman,R. (1972) Effects of comprehension on the retention of prose. *Journal of Experimental Psychology, 97,* 216-222.

Dosher, B. A., and Corbett, A. T. (1982). Instrument inferences and verb schemata. *Memory and Cognition, 10,* 531-539.

Dowty, D.R. (unpublished). Monotonicity-based logics and why natural languages have negative polarity and negative concord marking. Ohio State University.

Dowty, D.R., Peters , P.S, and Wall, R. (1981) *An introduction to Montague Grammar.* Dordecht: Reidel.

Ehrlich, K. (1980) The comprehension of pronouns. *Quarterly Journal of Experimental Psychology, 32,* 247-255.

Ehrlich, K., and Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of Verbal Learning and Verbal Behavior, 22,* 75-87.

Ehrlich, S. F., and Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20,* 641-655.

Fauconnier, G. (1979). Implication reversal in a natural language. In F. Guenther, and S.J. Schmidt (Eds.), *Formal semantics and pragmatics for natural language.* Dordecht: Reidel.

Fauconnier, G. (1985). *Mental Spaces: Aspects of meaning construction in natural language.* Cambridge: Cambridge University Press.

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., and Perry, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology. 20,* 400-409.

Fodor, J. A. (1983) *Modularity of mind.* Cambridge, MA: MIT press.

Fodor, J. D. (1982) The mental representation of quantifiers. In S. Peters and E. Saarinen (Eds.), *Processes, beliefs, and questions.* Dordecht: Reidel.

Fodor, J. D., Fodor, J. A., and Garrett, M. F. (1975). The psychological unreality of semantic representations. *Linguistic Inquiry, 6,* 515-531.

Frazier, L. (1995) On interpretation: Minimal 'lowering'. Paper presented at the 'Architectures and mechanisms for language processing' conference, December 1- 2, The Carlton Highland Hotel, Edinburgh.

Frazier, L., and Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6,* 291-325.

Frazier, L. and Rayner, K. (1982) Making and correcting errors during sentence comprehension: eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14,* 178-210.

Frege, G. (1892) 'Uber Sinn und Bedeuteng'. Translated in Geach, P. and Black, M. (1966) *Tranlations from the Philosophical writings of Gottlob Frege,* 56-78, Blackwell: Oxford.

Galambos, J. A., and Rips, L. J. (1982). Memory for routines. *Journal of Verbal Learning and Verbal Behavior, 21,* 260-281.

Garnham, A. (1979) Instantiation of verbs. *Quarterly Journal of Experimental Psychology, 31,* 207-214.

Garnham, A. (1981). Mental models as representations of text. *Memory and Cognition, 9,* 560-565.

Garnham, A. (1982). Testing psychological theories about inference making. *Memory and Cognition, 10,* 341-349.

Garnham, A. (1987). *Mental models as representations of discourse and text.* Chichester: Ellis Horwood.

Garnham, A. (1992). Minimilism versus constructionism: A false dichotomy in theories of inference during reading. PSYCOLOQUY 3(63) reading-inference-1.1

Garnham, A. (1993). An impartial view of inference making. PSYCOLOQUY 3(64) reading-inference-.11

Garnham, A., and Oakhill, J. (1985). On-line resolution of anaphoric pronouns: Effects of inference making and verb semantics. *British Journal of Psychology, 76,* 385-393.

Garnham, A., Oakhill, J., and Cruttenden, H. (1992). The role of implicit causality and gender cue in the interpretation of pronouns. *Language and Cognitive Processes, 7,* 231-255.

Garrod S. C., and Sanford, A. J. (1977). Interpreting anaphoric relations: the intergration of semantic information while reading. *Journal of Verbal Learning and Verbal Behavior, 16,* 77-90.

Garrod S. C., and Sanford, A. J. (1982) The mental representation of discourse in a focused memory system: Implications for the interpretation of anaphoric noun-phrases. *Journal of Semantics, 1,* 21-41.

Garrod S. C., and Sanford, A. J. (1985). On the real-time character of interpretation during reading. *Language and Cognitive Processes, 1,* 43-59.

Garrod S. C., and Sanford, A. J. (1991) Referential processing in reading: Focusing on roles and individuals. In D. A. Balota, G. B. Flores d'Arcais and K. Rayner (Eds). *Comprehension Processes in Reading,* 465-484.

Garrod S. C., Freudenthal, D., and Boyle, E. (1994). The role of different types of anaphor in the on-line resolution of sentences in a discourse. *Journal of Memory and Language, 32,* 39-68.

Garrod S. C., O'Brien, E. J., Morris, R. K,, and Rayner, K. (1990). Elaborative inferencing as an active or passive process. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 16, 250-257.

Garrod, S. C. (1985). Incremental pragmatic interpretation versus occasional inferencing during fluent reading. In G. Rickheit and H. Strohner (Eds.), *Inferences in text processing.* Elsevier: North-Holland.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition, 32,* 99-156.

Gernsbacher, M. A., and Robertson, R. R. W. (1992). Knowledge activation versus sentence mapping when representing fictional characters' emotional states. *Language and Cognitive Processes, 7,* 353-371.

Gil, D. (1982). Quantifier scope, linguistic variation and natural language semantics. *Linguistics and Philosophy, 5,* 421-472.

Gillen, K. (1991). The comprehension of doubly-quantified sentences. Unpublished doctoral thesis. University of Durham.

Glenberg, A. G., and Mathew, S. (1992). When minimalism is not enough: Mental models in reading comprehension. PSYCOLOQUY 3(64) reading-inference-2.1

Graesser, A. C., and Kreuz, R. J. (1993). A theory of inference generation during text comprehension. *Discourse Processes, 16,* 145-160.

Graesser, A. C., Singer, M., and Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101,* 371-395.

Greene, J. M. (1970). The semantic function of negatives and passives. *British Journal of Psychology, 61,* 17-22.

Greene, S. B., McKoon, G., and Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology, 18,* 266-283.

Grice, H. P. (1961). The causal theory of perception. *Proceedings of the Aristotelian Society, Supplementary vol. 35,* 121-152.

Grice, H. P. (1973) Probability, defeasibility and mood operators. Paper delivered at the Texas Conference on Perfomatives, Presuppositions and Implicatures.

Grimshaw, J. (1990). *Argument structure.* Cambridge, MA: MIT Press.

Hall-Partee, B. (1979). Semantics - Mathematics or psychology? In R. Bäurle, U. Egli, and A von Stechow (Eds.), *Semantics from different points of view.* Springer-Vertag: Berlin.

Halliday, M. A. K., and Hasan, R. (1976) *Cohesion in English.* London: Longman.

Haviland, S. E. and Clark, H. H. (1974) What's new? Aquiring new information as a process of comprehension. *Journal of Verbal Learning and Verbal Behavior, 13,* 512-521.

Hobbs, J., and Sheiber, S. (1987). An algorithm for generating quantifier scopings. *Computational Linguistics, 13,* 47-63.

Hörmann, H. (1983). The calculating listener or how many are einige, mehrere and ein paar (some, several and a few)? In R. Bäurle, C. Shwarze, and A. von Stechow (Eds.), *Meaning, use and the interpretation of language.* Walter de Gruyter: Berlin.

Inhoff, A. W., and Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception and Psychophysics, 40,* 431-439.

Ioup. G. (1978). Some universals for quantifier scope. In J. Kimball (Ed.) *Syntax and Semantics (vol. 4),* New York: Academic Press.

Jackendoff, R. (1972). *Semantic interpretation in generative grammar.* Cambridge, MA: MIT Press.

Jarvella, R. J., and Lundquist, L. (1994). Scales in the interpretation of words, sentences and texts. *Journal of Semantics,* 10, 171-198.

Johnson, M. K., Bransford, J. D. and Solomon, S. (1973). Memory for tacit implications of sentences. *Journal of Experimental Psychology, 98,* 203-205.

Johnson-Laird, P. N., and Steedman, M. (1978). The psychology of syllogisms, *Cognitive Psychology, 10,* 64-99.

Johnson-Laird, P. N., Byrne, R. M. J., and Tabossi, P. (1989). Reasoning by model: The case of multiple quantification. *Psychological Review, 96,* 658-673.

Johnson-Laird, P.N. (1969). On understanding logically complex sentences. *Quarterly Journal of Experimental Psychology, 21,* 1-13.

Johnson-Laird, P.N. (1982) Formal semantics and the psychology of meaning. In S. Peters and E. Saarinen (eds) *Processes, Beliefs and Questions.* Dordecht: Reidel.

Johnson-Laird, P.N. (1983) *Mental Models. Towards a cognitive science of language, inference and consciousness.* Cambridge University Press: Cambridge.

Johnson-Laird, P.N., and Byrne, R. (1991). *Deduction*. Hove: Lawrence Erlbaum Associates Ltd.

Just, M. A. (1974). Comprehending quantified sentences: The relation between sentence-picture and semantic memory verification. *Cognitive Psychology, 6,* 216-236.

Just, M. A., and Carpenter, P. A. (1971). Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior, 10,* 244-253.

Just, M. A., and Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329-354.

Katz, J. J. and Postal, P. M. (1965) *An integrated theory of linguistic descriptions.* Cambridge, MA: MIT Press.

Keenan, E. L., and Stavi, J. (1986). A semantic characterisation of natural language determiners, *Linguistics and Philosophy, 9,* 253-326.

Keenan, J. M., Baillet, S. D., and Brown, P. (1984). The effects of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior, 23,* 115-126.

Keenan, J. M., Golding, J. M., and Jennings, T. M. (1991). Which elaborative inferences are drawn during reading? A question of methodologies. In D. A. Balota, G. B. Flores d'Arcais, and Rayner, K, *Comprehension processes in reading.* Lawrence Erlbaum Associates: Hillsdale, NJ.

Kempson, R. M. (1977) *Semantics*. Cambridge: Cambridge University Press.

Kintsch, W. (1974) *The Representation of Meaning in Memory.* Hillsdale, NJ: Lawrence Erlbaum.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95,* 163-182.

Kintsch, W. (1993). Information accretion and reduction in text processing: Inferences. *Discourse Processes, 16,* 193-202.

Kintsch, W. and van Dijk, T. A. (1978) Towards a model of text comprehension and reproduction. *Psychological Review, 85,* 363-394.

Klein, H. (1994). Only (a) few. Unpublished manuscript. Faculty of Letters, University of Groningen.

Klima, E. S. (1964). Negation in English. In J. A. Fodor and J. J. Katz (Eds.), *The structure of language.* Englewood Cliffs, NJ: Prentice-Hall.

Kurtzman, H. S., and MacDonald, M. C. (1993). Resolution of quantifier scope ambiguities. *Cognition, 48,* 243-279.

Ladusaw, S. (1979). On the notion of affective in the analysis of negative polarity items. Paper presented to the 1979 Meeting of the Linguistics Society of America, Los Angeles.

Lakoff, G. (1971). On generative semantics. In D. Steinberg and L. Jakobovits (Eds.), *Semantics.* MIT Press: Cambridge, MA.

Lakoff, G. (1972). Linguistics and natural logic. In G. Harman and D. Davidson (Eds.), *Semantics for natural language.* Dordecht: Reidel.

Laport, L. (1994). Cognitive aspects of quantification. Unpublished manuscript. Dutch Institute, University of Groningen.

Lea R. B., O'Brien D. P., Fisch, S. M., Noveck, I. A., and Braine, M. D. S. (1990). Predicting propositional logic inferences in text comprehension. *Journal of Memory and Language, 29,* 361-387.

Levinson, S. C. (1982). *Pragmatics.* Cambridge: Cambridge University Press.

Link, G. (1983). The logical analysis of plurals and mass terms, a lattice-theoretical approach. In R. Baurle, C. Shwarze, and A. vam Stechow (Eds.), *Meaning, use and the interpretation of language.* Walter de Gruyter: Berlin.

Link, G. (1987). Generalised quantifiers and plurals. In P. Gardenfors (Ed.), *Generalised quantifiers: Linguistic and logical approaches*. D. Reidel: Dordecht.

Liversedge, S. P., and Pickering, M. J. (1995). Investigations into analyses of eye movements during reading. Paper presented at the Eighth European Conference on Eye movements, University of Derby, U.K., 6-9 September, 1995.

Loftus, E. F., and Palmer, J. P. (1972). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13*, 585-589.

Long, D. L., and Golding, J. M. (1993). Superordinate goal inferences: Are they automatically generated during comprehension? *Discourse Processes, 16*, 55-73.

Long, D. L., Golding, J. M., and Graesser, A. C. (1992). A test of the on-line status of goal-related inferences. *Journal of Memory and Language, 31*, 634-637.

Lønning, J. T. (1987). Collective readings of definite and indefinite noun-phrases. In P. Gardenfors (Ed.), *Generalised quantifiers: Linguistic and logical approaches*. D. Reidel: Dordecht.

MacDonald, M. C. (1994). Probabilisitic constraints and syntactic ambiguity resolution. *Cognitive Psychology*, 24, 56-98.

MacDonald, M. C., Pearlmutter, N. J., and Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676-703.

Mani, K. and Johnson-Laird, P. N. (1982) The mental representation of spatial descriptions. *Memory and Cognition*, 10, 181-187.

Marslen-Wilson, W., Tyler, L. K., and Koster, C. (1993). Integretative processes in utterance resolution. *Journal of Memory and Language*, 32, 647-666.

May, R. (1985). *Logical form.* MIT Press: Cambridge, MA.

McCawley, J. D. (1981). *Everything that linguists have always wanted to know about logic.* Basil Blackwell: Oxford.

McKoon, G., and Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior, 19,* 668-682.

McKoon, G., and Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12,* 82-91.

McKoon, G., and Ratcliff, R. (1990). Dimensions of inference. In A. C. Graesser and G. H. Bower (Eds.), *The psychology of learning and motivation, 25,* 313-328.

McKoon, G., and Ratcliff, R. (1991). Textual inferences: models and measures. In D. A. Balota, G. B. Flores d'Arcais and K. Rayner (eds). *Comprehension Processes in Reading,* 403-421.

McKoon, G., and Ratcliff, R. (1992). Inferences during reading. *Psychological Review, 99,* 440-466.

Miller, G. A., Isard, S. D. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior, 2,* 217-228.

Montague, R. (1970/1974) The proper treatment of quantification in ordinary English. In R. Thomason (Ed.), *Formal philosophy: Selected papers of Richard Montague.* Yale University Press: New Haven.

Mostowski, A. (1957). On a generalisation of quantifiers. *Fundamental Mathematics, 44,* 12-36.

Moxey, L. M., and Sanford, A. J. (1987). Quantifiers and focus. *Journal of Semantics, 5,* 188-206.

Moxey, L. M., and Sanford, A. J. (1993a). *Communicating Quantities: A psychological perspective.* Hove: Lawrence Erlbaum Associates.

Moxey, L. M., and Sanford, A. J. (1993b). Prior expectation and the interpretation of natural language quantifiers. *European Journal of Cognitive Psychology, 5,* 73-91.

Moxey, L. M., and Sanford, A. J., and Barton, S. B. (1990). Control of attentional focus by quantifiers. In K. J. Gilhooly, M. T. G. Keane, R. H. Logie, and G. Erdos (Eds.), *Lines of thinking, vol 1,* Wiley: Chichester.

Newstead, S. E. (1988). Quantifiers as fuzzy concepts. In T. Zetenyi (Ed.), *Fuzzy sets in psychology.* Elsevier-North Holland: Amsterdam.

Newstead, S. E. (1993). Do verbs act as implicit quantifiers? Paper presented at the International Conference on the Psychology of Language and Communication, University of Glasgow.

Newstead, S. E., and Collis, J. M. (1987). Context and the interpretation of quantifiers of frequency. *Ergonomics, 30,* 1447-1462.

Nicol, J., and Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research, 18,* 5-20.

O'Brien, E. J., Duffy, S. A., and Myers, J. L. (1986). Anaphoric inference during reading. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12,* 346-352.

O'Brien, E. J., Shank, D. M., Myers, J. L., and Rayner, K. (1988). Elaborative inferences during reading: Do they occur on-line? *Journal of Experimental Psychology: Learning, Memory and Cognition, 14,* 410-420.

Oakhill, J., Garnham, A., and Vonk, W. (1989). The on-line construction of discourse models. *Language and Cognitive Processes, 4,* S263-S286.

Paris, S. G. and Lindauer, B. K. (1976) The role of inference in children's comprehension and memory for sentences. *Cognitive Psychology, 8,* 217-227.

Paterson, K. B., and Edden, R. (in preparation). Anaphoric reference and quantifier scope ambiguity. Paper submitted to Discourse Anaphora and Anaphor Resolution Colloquium, Lancaster University, 17-18th July, 1996

Pepper, S., and Prytulak, L. S. (1974). Sometimes frequently means seldom: Context effects in the interpretation of quantitative expressions. *Journal of Research in Personality, 8,* 95-101.

Perfetti, C. A. (1994). Why inferences might be restricted. *Discourse Processes, 16,* 181-192.

Pickering, M. J., and Traxler, M. J. (submitted). Plausibility and recovery from garden paths.

Potts, G. R., Keenan, J. M., and Golding, J. M. (1988). Assessing the occurance of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language, 27,* 399-415.

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception, 8,* 21-30.

Rayner, K., and Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity and lexical ambiguity, *Memory and Cognition, 14,* 191-201.

Rayner, K., and Frazier, L. (1987). Parsing temporarily ambiguous complements. *Quarterly Journal of Experimental Psychology, 39A,* 657-673.

Rayner, K., and McConkie, G. W. (1976). What guides a reader's eye movements? *Vision Research, 16,* 829-837.

Rayner, K., and Pollatsek, A. (1989). *The psychology of reading.* Prentice-Hall: Englewood Cliffs, NJ.

Rayner, K., Well, A. D., Pollatsek, A., and Bertera, J. H. (1982). The availability of useful information to the right of fixation during reading. *Perception and Psychophysics, 31,* 537-550.

Rickheit, G., Schnotz, W., and Ströhner, H. (1985). The concept of inference in discourse comprehension. In G. Rickheit and H. Strohner (Eds.), *Inferences in text processing.* Elsevier: North Holland.

Sag, I. A., and Hankamer, J. (1984). Toward a theory of anaphoric processing. *Linguistics and Philosophy, 7,* 325-345.

Sanford, A. J, and Moxey, L. M. (1995). Aspects of coherence in written language: A psychological perspective. In T. Givòn and M. Gernsbacher (Eds.), *Coherence in spontaneous text,* J. Benjamins: Amsterdam.

Sanford, A. J. (1990). On the nature of text-driven inference. In D. A. Balota, G. B. Flores d'Arcais, and K. Rayner (Eds.), *Comprehension processes in reading.* Lawrence Erlbaum Associates: Hillsdale, N.J..

Sanford, A. J., and Garrod S. C. (1981). *Understanding written language: Explorations in comprehension beyond the sentence.* John Wiley and sons: Chichester.

Sanford, A. J., and Garrod S. C. (1989). What, when and how?: Questions of immediacy in anaphoric reference resolution. *Language and Cognitive Processes, 4*, 235-262.

Sanford, A. J., and Garrod S. C. (1991). Referential processes in reading. In D. A. Balota, G. B. Flores d'Arcais, and Rayner, K, *Comprehension processes in reading.* Lawrence Erlbaum Associates: Hillsdale, NJ.

Sanford, A. J., and Lockhart, F. E. (1990). Description types and method of conjoining as factors influencing plural anaphora: A confirmation study of focus. *Journal of Semantics, 7*, 365-378.

Sanford, A. J., and Moxey, L. M. (1991) Notes on plural reference and the scenario mapping principle in comprehension. Research Paper HCRC/RP-23, Human Communication Research Centre, Edinburgh University.

Sanford, A. J., Moxey, L. M., and Paterson, K. B. (1996). Attentional focusing with quantifiers in production and comprehension. *Memory and Cognition, 24*, 144-155.

Seigel, S., and Castellan, N. J. Jr. (1988). *Nonparametric statistics for the behavioral sciences.* McGraw-Hill: New York.

Shillcock, R. (1982). The on-line resolution of pronominal anaphora. *Language and Speech, 25*, 385-401.

Singer, M. (1979). Processes of inference in sentence encoding. *Memory and Cognition, 7*, 192-200.

Singer, M. and Ferreira, F. (1983). Inferring consequences in story comprehension. *Journal of Verbal Learning and Verbal Behavior, 22*, 437-448.

Speelman, C. P., and Kirsner, K. (1990). The representation of text-based and situation-based information in discourse comprehension. *Journal of Memory and Language, 29,* 119-132.

Stevenson, R. J., and Vitkovich, M. (1986). The comprehension of anaphoric relations. *Language and Speech, 29,* 335-357.

Tanenhaus, M. K., Caroll, J. M., and Bever, T. G. (1976). Sentence-Picture verification models as theories of sentence comprehension: A critique of Carpenter and Just. *Psychological Review, 83,* 310-317.

Trueswell, J. C., Tanenhaus, M. K., and Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic information in syntactic disambiguation, *Journal of Memory and Language, 33,* 285-318.

Turing, A. M. (1950) Computing machinery and intelligence. *Mind, 59,* 433-460.

Tyler, L. K. and Marslen-Wilson, W. D. (1982). The resolution of discourse anaphora: Some on-line studies. *Text,* 2, 263-291.

van Bentham, J. (1983). Determiners and logic. *Linguistics and Philosophy, 6,* 447-478.

van Bentham, J. (1986). *Essays in logical semantics.* D. Reidel: Dordecht.

van den Broek, P., Fletcher, C. R., and Risden, K. (1993). Investigations of inferential processes in reading. *Discourse Processes, 16,* 169-180.

van Eijk, J. (1986). Aspects of quanitifcation in natural language. Unpublished manuscript, Rijksunersiteit Groningen.

Vendler, Z. (1967) *Linguistics in Philosophy.* Cornell University Press: Ithaca, NY.

Vonk, W. (1984). Eye movements during comprehension of pronouns. In A Gale and F Johnson (eds.), *Theoretical and applied aspects of eye movement research.* Amsterdam: North-Holland.

Vonk, W. (1985). The immediacy of inferences in understanding pronouns. In G. Rickheit and H. Strohner (Eds.), *Inferences in text processing.* Amsterdam: North-Holland.

Vonk, W., and Noordman, L. (1991). On the control of inferences in text understanding. In D. A. Balota, G. B. Flores d'Arcais, and Rayner, K, *Comprehension processes in reading.* Lawrence Erlbaum Associates: Hillsdale, NJ.

Westerståhl, D. (1984). Determiners and context sets. In J. van Bentham and A. ter Meulen (Eds.), *Generalised quantifiers.* Foris: Dordecht.

Westerståhl, D. (1989). Quantifiers in formal and natural languages. In D. Gabbay, and F. Guenthner (Eds.), *Handbook of philosophical logic, 4,* 1-131.

Winer. B. J. (1971). *Statistical principles in experimental design.* McGraw-Hill: New York.

Yule, G. (1981). Interpreting anaphora without identifying reference. *Journal of Semantics, 1,* 315-322.

Zwaan, R. A., and Graesser, A. C. (1992). Reading goals and situation models. PSYCOLOQUY 3(64) reading-inference-.3

Zwarts, F. (to appear). Three types of polarity. In F. Hamm and E. Hinrichs (Eds.), *Semantics.*

Appendix 1:

Experimental and filler materials for Experiments 1 to 4

Experimental materials.

The following is a list of the experimental materials used in Experiments 1 to 4. The alternative forms for the quantifier and reference manipulations are listed in square brackets. These were quantified by *a few* and *few* in Experiments 1 & 3, and quantified by *many* and *not many* in Experiments 2 & 4. Experiments 2 & 4 were eye tracking studies in which the experimental materials were divided into regions text prior to analysis. The regions divisions are indicated by slashes.

Writing an essay.

The literature students were set an essay on Samuel Beckett and modern literary theory. / **[A few | Few]** of the students produced readable results. /Their **[clarity | confusion]** /demonstrated /how well they understood the topic.
Was the essay on Harold Pinter?

On an identity parade.

Some soldiers from the local barracks were put on an identity parade alongside the accused man. /**[A few | Few]** of the soldiers looked like the accused. /Their **[similarities | differences]** /were mentioned /by the victim.

At the funfair.

After having a ride on the big wheel a group of college students decided to try the coconut shy. /**[A few | Few]** of the students managed to hit a coconut. /Their **[accuracy | innacuracy]** /earned /them a commiseration prize.

In a hospital.

The doctor needed permission from some of the patients before she tested a new drug on them. /**[A few | Few]** of the patients agreed to act as guinea pigs. /Their **[consent | refusal]** /was noted /by the hospital registrar.

Testing job applicants.

Prospective air traffic controllers had to fill in a personality questionnaire then sit a series of aptitude tests. / **[A few | Few]** of the applicants passed. /Their **[success | failure]** /confirmed /the organiser's expectations.

A job offer.

Some accountancy trainees were offered a job with a major international company. /**[A few | Few]** of the accountants agreed to work for the company. /Their **[acceptance | rejection]** /was sent /by return of post.

In the court.

Some local youths were arrested during a police raid at a party and accused of drug-dealing. /**[A few | Few]** of the youths were found guilty of the crime. /Their **[aquittal | conviction]** /was /a relief to the whole neighbourhood/.

Coal mining.

The Coal Board wanted to increase productivity by making changes to working practices. /**[A few | Few]** of the miners believed the promises of job safety. /Their **[trust | distrust]** /was /the result of a previous deal.

At the gym.

Some weight-lifters from the gym competed to see who could lift a heavy dumb-bell. /**[A few | Few]** of the weight-lifters managed to lift it off the ground. /Their **[strength | weakness]** /surprised /their friends.

Getting exam results.

The girls crowded round the notice-board to see the examination results for chemistry and biology. /**[A few | Few]** of the girls did very well. /Their **[joy | grief]** /was /uncontrollable.

Running a marathon.

Some boys decided to enter the marathon. /**[A few | Few]** of the boys finished the race. /Their **[fitness | unfitness]** /surprised /their friends.

In the train station.

Passengers often had difficulty finding the platform for the evening train to Birmingham. /**[A few | Few]** of the staff were willing to give directions. /Their **[helpfulness | unhelpfulness]** /was reported /to the manager.

At the theatre.

At the end of the performance the actors came to the front of the stage and took a bow. /**[A few | Few]** of the audience expressed appreciation. /Their **[applause | silence]** /was /a surprise to the theatre manager.

In the office.

The insurance office had a new computer system installed over the Christmas holidays. /**[A few | Few]** of the secretaries had previously used the system. /Their **[experience | inexperience]** /made /the others feel more relaxed.

In the bar.

One of the barstaff liked to joke with female customers when they asked for a drink. /**[A few | Few]** of the woman appreciated his sense of humour. /Their **[pleasure | displeasure]** /was /obvious to everyone.

In a restaurant.

The office party went to an Italian restaurant for their Christmas dinner. /**[A few | Few]** of the office staff treated the waiter with any respect. /Their **[courtesy | rudeness]** /was reported /to the office manager.

Going to a party.

The girls asked their parents if they could go to an all-night party at a class-mate's house. /**[A few | Few]** of the parents allowed their daughter to go. /Their **[laxness | strictness]** /reflected /their own up-bringing.

A public meeting.

Local MP's were invited to take part in a public inquiry about proposals to build a nuclear power station. /**[A few | Few]** of the MP's attended the meeting. /Their **[presence | absence]** /helped /the meeting run more smoothly.

At a boy-scout camp.

It was dark before the boy-scouts finished putting up the tents and collected wood for a camp-fire. /**[A few | Few]** of the scouts went to bed immediately. /Their **[wakefulness | sleepiness]** /annoyed /the scoutmaster.

In a classroom.

The trainee mathematics teacher was nervous about giving her first unsupervised lesson in such a rough school. /**[A few | Few]** of the class misbehaved. /Their **[quietness | noisiness]** /was reported /to the real teacher.

At the football match.

The stadium was packed to see Falkirk play East Stirling in the final of the Stirlingshire Cup. / **[A few | Few]** of the Falkirk fans expected Falkirk to win. /Their **[optimism | pessimism]**/ was rewarded/ by an early goal.

At a church meeting.

The evangelist told the audience that he would cure the sick and disabled among them. /**[A few | Few]** of the audience thought that he could perform miracles. /Their **[belief | disbelief]** /was shared /by the church authorities.

At the swimming pool.

A class of school children had been taking swimming lessons for the last couple of weeks. /**[A few | Few]** of the children could manage the back-stroke. /Their **[competence | incompetence]** /surprised /the instructor.

Caught shop-lifting.

A group of boys were suspected of stealing magazines and sweets from the newsagents on the corner. / **[A few | Few]** of the boys admitted to the crime. /Their **[honesty | dishonesty]** /affected /the punishment they received.

Filler materials for Experiment 1 & 3

The folowing materials were used as filler items in Experiment 1. Some of these were materials for another experiment. The alternative forms for that experimental manipulation are displayed in brackets.

In a taxi

A taxi-driver dropped a man and a woman off at the station. On the way they had run into a traffic jam. The taxi-driver and the woman had been arguing but the man had remained calm. The taxi-driver told the woman that he had been arguing with that she wouldn't miss the train.

Studying for exams

A student had been working hard for her exams. On the way to the library she met two boys from her tutorial group. The student and one of the boys had been studying together but the other never seemed to work. The student told the boy that had been studying with her that his friend was a bad influence.

Crossing the channel.

A coach party was booked on the hovercraft from Dover to Calais. On arriving at Dover the coach passengers were annoyed to find that the hovercraft was cancelled due to poor weather. After complaining to the hovercraft company they took a ferry. The [ferry was | ferries were] three hours late to Calais and they missed their connection.

Going to the opera.

Alison and her friends were given free tickets for the opera at Covent Garden. The girls were excited about going, but at the same time worried that they would look scruffy compared to the rest of the audience. Before the show they hired an expensive dress. The [dress was | dresses were] made from pure silk.

In a department store.

The department store employed a team of plain-clothes security guards to protect the stock. The security guards walked around the store and kept in contact with each other using walkie-talkies. After only an hour on the job they caught a shop-lifter. The [shop-lifter was | shop-lifters were] handed over to the police.

At a school.

The headmaster had reached retirement age and the school needed to replace him. Most of the teachers at the school were keen to get the job since it meant a much larger salary. As soon as the post was advertised they requested an application form. The [application form was | application forms were] sent by return of post.

At the Post Office.

There was a crowd of people waiting for the Post Office to open. When it was opened there was only one assistant serving behind the counter. To get served quickly they formed a queue. The [queue was | queues were] stretched to the door.

At the Booker Prize.

It was well-known that the authors short-listed for the

Booker Prize hated each other. However the authors had avoided making public comments before the final decision had been reached. After the awards ceremony they allowed an interview. The [interview was | interviews were] highly critical of the competition.

In the hospital

After the surgeon examined the girl with the broken leg, he decided he would have to take immediate action. He he'd had a lot of experience with serious injuries and knew what to do next. He quickly injected the girl with a painkiller.

At the airport

A film star arrived at Heathrow and was surrounded by journalists. They noticed two women who seemed to be with him. The film star and one of the women had been travelling together but the other woman was a friend who had come to meet them. The film star told the woman that he had been travelling with that the journalists were a real drag.

The union meeting.

The factory workers were furious when the management refused to give a wage rise. They organised a meeting to call for some form of action. After a short debate they passed a strike motion. The [motion was | motions were] to take effect from the next day.

Watching a tennis match.

The Wimbledon crowd were prepared for bad weather. The forecast was for heavy thunderstorms so they were dressed in waterproofs. As soon as it started to rain they put up an umbrella. The [umbrella was | umbrellas were] not enough to keep them dry.

Robbing a bank.

The burglars had broken into the bank in the early hours of the morning. They smashed open the safe and stole several thousand pounds. Before leaving the bank they were recorded on a video. The [video was | video s were] produced as evidence in court.

Arresting terrorist suspects.

After the terrorist bombing the police rounded up a number of suspects. The suspects were taken to separate police stations and questioned for hours. After a while they made a confession. The [confession was | confessions were] later retracted in court.

In parliament.

The MP's were holding a debate on the state of the economy. The evasiveness of the government spokepersons was annoying the MP's. Eventually it became too much and they heckled a speaker. The [speaker was | speakers were] forced to sit down.

Visiting grandparents.

Mr Smith always offered his grandchildren some fruit when they visited him. Today the grandchildren could choose between apples and pears. After looking the fruit over they took an apple. The [apple was | apples were] sour and made the children feel sick.

At the library

John was being teased by two female friends for working too hard in the library. After a few really cutting remarks he got very angry. John hit the girl with a book.

In the street

A burglar looked over most of the houses in the street. Several of them had lights on so he concentrated on the others. Eventually he picked his target and broke in by the rear window.

Collecting for charity.

An Oxfam supporter was collecting money for development work in Bangladesh. He stopped passing shoppers and asked them to make a donation. After giving some money they signed a petition. The [petition was | petitions were] calling for an increase in Third World aid.

At a Youth Hostel.

As the weather was very bad the boy-scouts had decided to stay at a nearby Youth Hostel for the night. The boys were shown to a large communal dormitory. Before unpacking they chose a bed. The [bed was | beds were] hard and uncomfortable.

Glue-sniffing.

A group of glue-sniffers usually hung around the stairwell of the block of flats. They annoyed the residents who reported them to the police. The glue-sniffers decided to take revenge. After piling newspapers in the stairwell, they started a fire. The [fire was | fires were] put out by one of the residents.

Caught by a policeman.

A policeman caught some boys writing graffiti on a wall. He asked them for their names and addresses. After a pause they gave a false name. The [name was | names were] not good enough to fool the policeman and he arrested them.

At the cinema.

The girls agreed to go to the cinema but could not decide which film to watch. One of them liked horror films, while another preferred adventure films. After an argument they decided on a comedy. The [comedy was | comedies were] showing at the Odeon cinema.

At a party.

Everyone at the party had to provide some entertainment. The more musical guests played guitar or sang. Some of the boys were too embarrassed to sing. As an alternative they told a joke. The [joke was | jokes were] appreciated by the other guests.

On a train

John was sitting next to an old woman in the train. After an hour the woman fell off her chair and lay on the ground moaning in pain. John panicked and pulled the emergency cord. The train came to a halt.

At the village hall

A woman was giving dancing lessons in the village hall. Among her pupils were two builders from the village. The woman and one of the builder danced together but the other just sat and watched. The woman told the builder that his footwork needed more practice.

After the concert.

After the concert the band packed their equipment into a van. They found it easiest to pack the light instruments into the van first. After packing the guitars into the van they carried a piano. The [piano was | pianos were] the last instrument on board.

At a funeral.

By coincidence all of John's uncles died in the same week. They were given separate funerals in different parts of the country, and John attended them all. It turned out that they had left a will. The [will was | wills were] very generous to John.

On a building site.

The building contractors had hired some plasterers too early. To cover up the mistake the bricklayers were persuaded to speed up their own work. Before the end of the shift they finished a house. The [house was | houses were] enough to keep the plasterers busy for a while.

In a school.

The new school principal was very strict and introduced a number of new rules. Girls were not allowed to wear make-up or jewellery. No-one was happy with the new rules, and the sixth year girls were furious. As an act of rebellion they wore a necklace. The [necklace was | necklaces were] enough to have them suspended.

A saxophone quartet.

The saxophone quartet had spent some time touring the United States. Their performances were acclaimed and they were tipped for the big-time. After a few months rest they recorded an album. The [album was | albums were] appreciated by the critics.

On a summer holiday.

During the summer Tom's friends travelled all over the world. One went to Australia, one went to Canada while others visited different parts of Europe. Before coming home they sent him a postcard. The [postcard was | postcards were] enough to make Tom feel envious.

Filler materials for Experiments 2 & 4

The following materials were used as filler items in Experiment 2. Some of these were materials for another experiment. The alternative forms for that experimental manipulation are displayed in brackets.

In a taxi.

A taxi-driver dropped a man and a woman off at the station. On the way they had run into a traffic jam. The taxi-driver and the woman had been arguing but the man had remained calm. The taxi-driver told the woman that he had been arguing with that she wouldn't miss the train.

Studying for exams.

A student had been working hard for her exams. On the way to the library she met two boys from her tutorial group. The student and one of the boys had been studying together but the other never seemed to work. The student told the boy that had been studying with her that his friend was a bad influence.

Crossing the channel.

A coach party was booked on the hovercraft from Dover to Calais. On arriving at Dover the coach passengers were annoyed to find that the hovercraft was cancelled due to poor weather. After complaining to the hovercraft company, they took a ferry. Their [ferry was | ferries were] the last one(s) to leave from Dover that day.

Going to the opera.

Alison and her friends were given free tickets for the opera at Covent Garden. The girls were excited about going, but at the same time worried that they would look scruffy compared to the rest of the audience. Before the show they hired an expensive dress. The [dress was | dresses were] the prettiest one(s) that the girls had ever worn in their lives.

In a department store.

The department store employed a team of plain-clothes security guards to protect the stock. The security guards walked around the store and kept in contact with each other using walkie-talkies. After only an hour on the job, they caught a shoplifter. Their [shop-lifter was | shop-lifters were] the only one(s) that the department store caught that day.

At a school.

The headmaster had reached retirement age and the school needed to replace him. Most of the teachers at the school were keen to get the job since it meant a much larger salary. As soon as the post was advertised they requested an application form. Their [application form was | application forms were] the only one(s) that the committee considered for the post.

At the Post Office.

There was a crowd of people waiting for the Post Office to open. When it was opened there was only one assistant serving behind the counter. To get served quickly, they formed a queue. Their [queue was | queues were] the slowest one(s) that the customers had ever waited in.

The Booker Prize.

It was well-known that the authors short-listed for the Booker Prize hated each other. However the authors had avoided making public comments before the final decision had been reached. After the awards ceremony they gave an interview. The [interview was | interviews were] the only one(s) that the authors gave on the subject.

In the hospital.

After the surgeon examined the girl with the broken leg, he decided he would have to take immediate action. He'd had a lot of experience with serious injuries and knew what to do next. He quickly injected the girl with a painkiller.

At the airport.

A film star arrived at Heathrow and was surrounded by journalists. They noticed two women who seemed to be with him. The film star and one of the women had been travelling together but the other woman was a friend who had come to meet them. The film star told the woman that he had been travelling with that the journalists were a real drag.

The union meeting.

A group of factory workers were furious when the management refused to give a wage rise. They organised a union meeting to call for some form of action. After a short debate, they passed a strike motion. The [strike motion was | strike motions were] the first one(s) that the union had passed in years.

Watching a tennis match.

The Wimbledon crowd were prepared for bad weather. The forecast was for heavy thunderstorms so they were dressed in waterproofs. As soon as it started to rain they put up an umbrella. The [umbrella was | umbrellas were] the biggest one(s) that the officials allowed into the court.

Robbing a bank.

The burglars had broken into the bank in the early hours of the morning. They smashed open the safe and stole several thousand pounds. Before leaving the bank, they were recorded on video. The [video recording was | video recording were} the best clue that the police had to their identity

Arresting terrorist suspects.

After the terrorist bombing the police rounded up a number of suspects. The suspects were taken to separate police stations and questioned for hours. After a while they gave a confession. Their [confession was | confessions were] the unlikeliest one(s) that the judge had heard in his career.

In parliament.

The MP's were holding a debate on the state of the economy. The evasiveness of the government spokespersons was annoying the MP's. When it became too much they heckled a speaker. Their [target was | targets were] the dullest speakers in the debate that day.

Visiting grandparents.

Mr Smith always offered his grandchildren some fruit when they visited him. Today the grandchildren could choose between apples and pears. After looking the fruit over they took an apple. Their [apple was | apples were] the juiciest one(s) that the children had tasted in weeks.

At the library.

John was being teased by two female friends for working too hard in the library. After a few really cutting remarks he got very angry. John hit the girl with a book.

In the street.

A burglar looked over most of the houses in the street. Several of them had lights on so he concentrated on the others. Eventually he picked his target and broke in by the rear window.

Collecting for charity.

Some Oxfam supporters were collecting money for development work in Bangladesh. They stopped passing shoppers and asked them to make a donation. After giving some money, they signed a petition. The [petition was | petitions were] the largest one(s) that Oxfam had organised for many years.

At a Youth Hostel.

As the weather was very bad the boy-scouts had decided to stay at a nearby Youth Hostel for the night. The boys were shown to a large communal dormitory. Before unpacking they chose a bed. The [bed was | bed were] the hardest one(s) that the boys had ever slept on in their lives.

Glue-sniffing.

A group of glue-sniffers usually hung around the stairwell of the block of flats. They annoyed the residents who reported them to the police. The glue-sniffers decided to take revenge. After piling newspapers in the stairwell, they started a fire. The [fire was | fires were] the worst one(s) that the flats had suffered for many years.

Caught by a policeman.

A policeman caught some boys spraying graffiti on a wall. He took out his notebook and asked the boys for their names and addresses. After thinking for a moment they gave a false name. The [false name was | false names were] the worst one(s) that the policeman had heard in his career.

At the cinema.

The girls decided to go to the cinema but could not decide which film they wanted to watch. One of them liked science fiction and another preferred adventure films. After a heated argument they decided on a comedy. The [film was | films were] the best one(s) that the girls had seen in months.

At a party.

Everyone at the party had to provide some entertainment. The more musical guests played guitar or sang. Some of the boys were too embarrassed to sing. As an alternative they told a joke. Their [joke was | jokes were] the worst one(s) that the other guests had heard in a long time.

On a train.

John was sitting next to an old woman in the train. After an hour the woman fell off her chair and lay on the ground moaning in pain. John panicked and pulled the emergency cord. The train came to a halt.

At the village hall.

A woman was giving dancing lessons in the village hall. Among her pupils were two builders from the village. The woman and one of the builder danced together but the other just sat and watched. The woman told the builder that his footwork needed more practice.

After the concert.

After the concert the musicians packed their equipment into a van. They found it easiest to pack the light instruments into the van first. After packing the guitars, they carried a piano. The [piano was | pianos were] the last instrument(s) that the musicians loaded into the van.

At a funeral.

By coincidence several of John's uncles died in the same week. They were given separate funerals in different parts of the country, and John attended them all. It turned out that they had left a will. The [will was | wills were] the most generous one(s) that Tom had heard in his life.

On a building site.

The building contractors had hired some plasterers too early. To cover up the mistake the bricklayers were persuaded to speed up their own work. Before the end of the shift they finished a house. Their [house was | houses were] the first one(s) that the plasterers tackled that week.

In a school.

The new school principal was very strict and introduced a number of new rules. Girls were not allowed to wear make-up or jewellery. No-one was happy with the new rules, and the sixth year girls were furious. As an act of rebellion they wore a necklace. Their [necklace was | necklace were] the first thing that the principal saw that morning.

A saxophone quartet.

The saxophone quartet had spent some time touring the United States. Their performances were acclaimed and they were tipped for the big-time. After a few months rest, they recorded an album. Their [album was | albums were] the best one(s) that the critics had heard for many years.

On a summer holiday.

During the summer Tom's friends travelled all over the world. One went to Australia, one went to Canada while others visited different parts of Europe. Before coming home they sent a postcard. Their [postcard was | postcards were] the only one(s) that Tom received that summer.

Appendix 2:

Materials for Experiment 5

Experimental materials

The following are the set of experimental materials used in Experiment 5, with their region of analysis divisions. The alternative quantifier and reference manipulations are highlighted, and the critical anaphoric region is italicised. The materials were double lined-spaced in the presentation.

At the boy-scout camp

It was late before the boy-scouts had finished setting-up camp. /**[A few | Few]** of the boys were **[wide awake |very tired]** after the work, /so /*they slept* /*soundly* /until /morning /arrived.

In a retirement home

All of the residents were usually in the TV room during Brookside. /**[A few | Few]** of the residents were **[excited | bored]** with the plot, /so /*they watched* /*intently* /when /the programme /began.
Was Coronation Street on the TV?

Caught shop-lifting

A group of school-boys were suspected of stealing from the local shop. / **[A few | Few]** of the boys were **[innocent | guilty]** of the crime,/so /*they confessed* /*immediately* /when /the police /arrived.

At the theatre

The local theatre company put on a controversial work.
/**[A few | Few]** of the audience were **[pleased by | disgusted]** by the performance/, so /*they applauded* /*appreciatively* /when /the play /finished.

At a conference

The conference delegates used the university cafeteria for lunch. /**[A few | Few]** of the delegates were **[impressed | disappointed]** with the food/, so /*they complained* /*bitterly* /when /a manager /arrived.
Did the delegates use the university cafeteria?

Keeping fit

The security guards were almost all unfit and overweight.
/**[A few | Few]** of the men were **[enthusiastic | apathetic]** about improving their health/, so /*they exercised* /*vigorously* /before /the shift started.

In a hospital

Some medical students had the chance to participate in a drugs trial. /**[A few | Few]** of the students were **[keen | reluctant]** to take part,/ so /*they refused* /*pointedly* /when /the medical authorities /asked.

Going to a party

Some of the school-boys asked their parents for beer to take to the party. /**[A few | Few]** of the parents **[allowed | forbade]** underage drinking,/ so /*they agreed* /*willingly* /when /the boys /asked.

In the classroom

The teacher lectured the entire class about behaving properly. / **[A few | Few]** of the children were **[praised | scolded]** by the teacher, / so / *they wept* / *tearfully* / until / their parents / arrived.

Visiting grandparents

The grandfather always told the children about his time in the navy. / **[A few | Few]** of the children were **[interested in | tired of]** the stories, / so / *they listened* / *carefully* / until / their bedtime / came.
Was the grandfather in the navy?

Sitting exams

The students found that a hectic social life interfered with their studies. / **[A few | Few]** of the students **[attended to | ignored]** their work, / so / *they failed* / *miserably* / when / the exams / were marked.

In the pub

Some men were playing on the fruit machine in the pub. / **[A few | Few]** of the men **[won | lost]** a lot of money,/ so / *they celebrated* / *extravagantly* / until / the bar / closed.

Catching a bus

As usual the bus was busy with commuters during rush hour. / **[A few | Few]** of the commuters **[managed | failed]** to get a seat, so / *they stood* / *awkwardly* / until / the journey / was over.
Did the commuters travel by train?

Catching flu.

The school was warned that a lot of children might fall prey to a flu epidemic. /**[A few | Few]** of the children **[caught | avoided]** the virus,/ so /*they sneezed* /*constantly* /until /the infection /had passed.

At the protest march

The animal rights demonstrators were badly treated by the police. /**[A few | Few]** of the demonstrators **[supported | condemned]** the police action, /so /*they protested* /*vociferously* /until /an apology /was given.

After the accident

The police blamed dangerous drivers for a major motorway pile-up. /**[A few | Few]** of the drivers **[admitted | denied]** responsibility, so /*they apologised* /*profusely* /when /the police /arrived.

Catching a train

Commuters were always in a hurry to catch the London train. /**[A few | Few]** of the commuters were **[early | late]** for the train,/ so /*they sprinted* /*frantically* /before /the train /departed.

In the university

A number of university staff were offered early retirement. /**[A few | Few]** of the staff **[accepted | rejected]** the offer,/ so /*they retired* /*gracefully* /when /the academic year /finished.

At the school football match

Some boys from the local team were playing against a nearby school. /**[A few | Few]** of the boys were **[friendly | hostile]** to the other team, /so /*they argued* /*incessantly* /until /the game /was abandoned.

After a shipwreck

Most of the passengers dived overboard when the ship sank. /**[A few | Few]** of the passengers **[managed | struggled]** to stay afloat, /so /*they waited* /*calmly* /until /the rescue services /arrived.
Did most of the passengers dive overboard?

At the casino

A group of men won a lot of money on the roulette wheel. / **[A few | Few]** of the men were **[careful | careless]** with their winnings, /so /*they gambled* /*recklessly* /until /the money /was gone.
Were the men playing poker?

In parliament

MPs were scheduled to debate increased taxation for gas and electricity. /**[A few | Few]** of the MPs were **[present | absent]** from the debate, /so /*they voted* /*unanimously* /after /the motion /was debated.

In the bar

One of the barstaff always made sexist remarks to female customers. / **[A few | Few]** of the women **[appreciated | detested]** his sense of humour, /so /*they objected* /*vocally* /when /the manager /arrived.

After the election

A crowd of people gathered to hear the election
result. / **[A few | Few]** of the crowd felt **[glad | sad]** about the outcome,/
so / *they clapped* / *enthusiastically* / when / the winner / spoke.

At playschool

The children all wanted to play with one of the toys at the
same time. / **[A few | Few]** of the children were **[passive | aggressive]** by
nature, / so / *they fought* / *violently* / until / the toy / broke.

After a hijacking

The hijackers were trapped in the aircraft and surrounded by
police. / **[A few | Few]** of the hijackers wanted to **[continue | end]** the siege,
/ so / *they surrendered* / *unconditionally* / before / the police / attacked.

At the demonstration

The local council wanted to build a new road through the
city. / **[A few | Few]** of the local residents **[approved | opposed]** the new
road, / so / *they demonstrated* / *noisily* / when / the workmen / arrived.
Did the local council want to build a road?

At the trial

Some Irishmen living in London were accused of terrorist
activities. / **[A few | Few]** of the Irishmen were **[aquitted | convicted]** at the
trial, / so / *they rejoiced* / *triumphantly* / when / they / were released.
Were the Irishmen living in Belfast?

A student house

A group of students moved into a new house.
/**[A few | Few]** of students felt **[warm | cold]** in the house,/
so / *they shivered* / *icily* / until / the heating / came on.

Begging in the street

A tramp was pestering shoppers to give him some
money. / **[A few | Few]** of the shoppers **[offered | refused]** to help,/
so / *they donated* / *generously* / when / the man / approached.

On holiday abroad

A group of tourists drank untreated tap water in the hotel.
/ **[A few | Few]** of the tourists **[suffered | avoided]** a bout of food poisoning,
/ so / *they vomited* / *copiously* / until / the doctor / arrived.

After a robbery

Local residents had suffered a spate of robberies.
/ **[A few | Few]** of the residents **[agreed | declined]** to join the neighbourhood
watch, / so / *they patrolled* / *regularly* / until / the thefts / stopped.

Filler materials

The following are the set of filler materials. Many of these were experimental materials for another study. Their alternative forms are listed in brackets.

The emergency crew asked [who would treat the accident victim. | where to treat the accident victim.] Skilfully the victim was treated by the [eager nurse | grass verge] immediately.

The head gardener decided [who should plant the shrubs. | where to plant the shrubs.] In the end the shrubs were planted by the [apprentice | greenhouse] that morning.

The lawyer explained [who had found the lost document. | where to find the lost document.] It seemed that the file was found by the [personnel manager | shredding machine] last night.

The lecturer did not know [who would find his lost wallet. | where to find his lost wallet.] Luckily the wallet was discovered by the [student teachers | theatre entrance] last night.

The general announced [who would carry out the ambush. | where to carry out the ambush.] As planned the convoy was ambushed by the [commando officers | deserted building} the next day.

The student knew [who had performed the burial of the man. | where the burial of the man took place.] Without doubt, the man was buried by the {bishop}/{chapel } twenty years ago.

The teacher decided [who should hide the prize. | where to hide the prize. After some thought the prize was hidden by the [intelligent pupil | grandfather clock] that afternoon.

The girl heard [who had discovered the body. | where the discovery of the body occurred.] Sadly, the body was discovered by the [grieving cousin | isolated stable] that afternoon.

The manager decided [who should arrest the shoplifter. | where to arrest the shoplifter.] Discreetly the shoplifter was arrested by the [security guard | changing rooms] very promptly.

The policeman wanted to know [who had carried out the stabbing | where the stabbing took place.] Apparently, the youngster was stabbed by the [convicts | tenement] last week.

The mafia had decided [who was to detonate the bomb. | where to detonate the bomb.] As intended, the bomb was detonated by the [nervous gangster | railway platform] during the night.

The driver asked [who would unload the delivery van. | where to unload the delivery van.] In fact, the van was unloaded by the [bored assistant | empty warehouse] very quickly.

The newscaster described [who had carried out the attacks. | where the attacks took place.] In fact the crimes were committed by the [sadistic mugger | desolate bridge] during the night.

The hotel staff wondered [who would find the lost girl. | where to find the lost girl.] In the end the child was discovered by the [manageress | escalator] during the evening.

The florist enquired [who had sold the last of the flowers. | where to sell the last of the flowers.] In fact, the flowers were sold by the [teenager | roadside] during the morning.

The producer announced [who would conduct the interview. | where to conduct the interview.] As usual, the athlete was interviewed by the [anxious youth | running track] after the final.

The farmer asked [who had pitched the big tent. | where to pitch the big tent.] Surprisingly, the tent was pitched by the [girl guide | deep river last month.

The police identified [what had caused the accident. | where the accident happened.] It turned out that the cyclist was knocked down by the [juggernaut | roundabout] that morning.

The owner pointed out [who had trained the horse. | where to train the horse.] Unusually, the horse was trained by the [proud jockeys | muddy paddock] last year.

The admiral decided [who should moor the ship. | where to moor the ship. As ordered, the ship was moored by the {captain}/{harbour} straight away.

The cop noted [who assaulted the foreman. | where the assault of the foreman occurred.] In fact, the foreman was assaulted by the [tattooed miner | colliery gates] that evening.

The police wondered [who would find the murder victims. | where to find the murder victims.] At last the bodies were dug up by the [gardeners | riverside] the next day.

The duke decided [who should set up the huge marquee. | where to set up the huge marquee.] As ordered, the marquee was set up by the [attendant | fountains] during the morning.

The supervisor asked [who would refuel the jumbo jet. | where to refuel the jumbo jet.] Finally the plane was refuelled by the [tired pilots | empty hangar] late on Friday.

The fact that the meat had gone down in price surprised the butcher
who was struggling to make ends meet.

Singing aloud always embarrassed the man.
He would mime to hymns in church in future.

The four weeks running up to Christmas are always hectic.

Jessica enjoyed a good bedtime story.
Within moments of her dad reading, she fell asleep.

The mechanic took a spanner from his toolbox.
It would be easy to fix the motorbike.

Every day the bus was about ten minutes late.
This made the passengers anxious and irritable.

The lion let out a huge roar to the hunter.
It was very protective of its young.

The loud music was not very appealing to the girl
who hated parties.

Leaves on the track caused the angry businessman to be delayed
for his important meeting.

The black Labrador enjoyed chasing after large wooden sticks
which his owner often threw for him.

The baker took the loaf out of the oven.
He would enjoy eating it for his lunch.

Appendix 3:

Materials for Experiments 6 & 7

Materials for Experiment 6

The following are the set of materials used in Experiment 6. These were presented to subjects in one of four continuation conditions. Either the plural pronoun began a new sentence, or was conjoined to the quantified sentence by *and, but* or *because.* The materials were also presented in one of two forms, where the quantified sentence either matched or violated pragmatic norms. These alternative forms are highlighted.

In the mathematics tutorial class.

At the end of term, all of the post-graduate mathematics students tried to answer an 'O-grade' exam paper for some fun. Only a few of the post-graduates were able to answer **[more | less]** than three questions. They /and they /but they / because they . . .

Working in a supermarket.

The supermarket desperately needed extra staff during the weeks before Christmas, so the manager hired some students. Only a few of the students were **[on time | late]** each morning. They /and they /but they / because they . . .

Getting exam results.

The girls crowded around the notice-board to see the exam results for chemistry and biology. Only a few of the girls scored more than **[ninety | five]** percent. They /and they /but they / because they . . .

At the graduation ceremony.

The University's graduation ceremony was always an extremely expensive affair. Only a few of the students could afford to hire **[graduation robes at seven pounds a day | a Rolls-Royce at one hundred pounds a day]**. They /and they /but they / because they . . .

In a restaurant.

The office party went to an up-market Italian restaurant for their Christmas dinner. Only a few of the office staff **[offered to buy the waiter a drink | ate with their mouths closed]**. They /and they /but they / because they . . .

Materials for Experiment 7

Experimental materials

The following are the set of experimental materials used in Experiment 7. The alternative quantifier and reference manipulations are highlighted. The materials were double lined-spaced in the presentation.

At the gym.

Some expert weight-lifters from the gym held a weight-lifting competition. Only a few of the weight-lifters managed to lift the **[heaviest | lightest]** dumbbell. Their **[strength | weakness]** was a surprise to their friends.
Was the competition held in a gym?

In the mathematics tutorial class.

At the end of term, all of the post-graduate mathematics students tried to answer an 'O-grade' exam paper for some fun. Only a few of the post-graduates were able to answer **[more | less]** than three questions. Their **[competence | incompetence]** was a surprise to the rest of the class.

Working in a supermarket.

The supermarket desperately needed extra staff during the weeks before Christmas, so the manager hired some students. Only a few of the students were **[on time | late]** each morning. Their **[enthusiasm | laziness]** was noted by the manager.

At the funfair.

After having a ride on the big wheel a group of college students decided to try the coconut shy. Only a few of the students managed to **[hit all three coconuts | could throw in the right direction]**. Their **[accuracy | inaccuracy]** was a surprise to their friends.
Had the students been on the rollercoaster?

At the train station.

Clapham Junction was busy all day with commuters waiting for connecting trains into London. Only a few of the trains arrived **[ahead of schedule | on time]**. Their **[early | late]** arrival annoyed the station master.

At a party.

Graeme hired the local club for his stag-night party. He invited all of his rugby club friends, and felt sure that it would be a wild night. Only a few of the guests **[were still standing by the end of the night | insisted on drinking orange juice all night]**. Their **[sobriety | drunkeness]** disappointed Graeme. Was it Graeme's stag night party?

Comparing heights.

All the women in the office decided to compare height using a tape measure. Only a few of the women were over **[four | six]** feet tall. Their **[tallness | shortness]** amused the others.

A health inspection.

The **[athletic club | cancer ward]** carried out a routine health inspection of its patients. Only a few of the **[patients | members]** were declared fit and well. Their **[good | poor]** health surprised the health inspector.

In the office.

The insurance office had a computer system installed over the Christmas break. Only a few of the secretaries had **[previously used this particular system | seen a computer]**. Their **[experience | inexperience]** was noted by the office manager.
Were the computers installed over Easter?

Running a marathon.

Some boys from a local college decided to enter the London marathon. Only a few of the boys finished the race with **[a reasonable time | record-breaking time]**. Their **[fitness | unfitness]** was a surprise to their friends.

At the theatre.

At the end of the performance the actors came to the front of the stage and took a bow. Only a few of the audience **[showed their appreciation | gave a long standing ovation]**. Their **[applause | silence]** was noted by the theatre critics.

Caught by a policeman.

A policeman shouted to some boys to stop spraying graffiti on a wall. Only a few of the boys **[paid attention to him and stopped | called him sir and apologised]**. Their **[respect | disrespect]** amused the policeman.

Buying CD's.

The boys went to John Menzies to buy some compact discs. Only a few of the CD's cost less than **[three | thirty]** pounds. Their **[expensiveness | cheapness]** surprised the boys.
Did the boys go to HMV?

In the train station.

Passengers often had difficulty finding the correct platform for the evening train to Birmingham. Only a few of the staff **[were willing to listen to passengers | would escort the passengers to the platform]** questions. Their **[helpfulness | unhelpfulness]** was reported to their manager.
Did the passenger want the Birmingham train?

At the swimming pool.

A class of school children had been taking swimming lessons for the last three years. Only a few of the children could manage **[a breadth of the pool | thirty lengths of the pool]**. Their **[aptitude | inaptitude]** surprised the instructor.

At the circus.

The teacher took her primary school class to see the lions and acrobats at the circus. Only a few of the children were **[terrified by the performance | stayed awake through the performance]**. Their **[excitement | boredom]** surprised the teacher.
Did the class see any acrobats?

Applying for a job.

It was very easy to get manual jobs in the area, so a group of school-leavers decided to apply for jobs as **[astronauts | labourers]**. Only a few of the school-leavers were successful in their application. Their **[success | failure]** surprised their friends.

At the boxing match.

Before the heavy weight bout, all of the boxers were weighed by a doctor. Only a few of the boxers were under **[fifteen | seven]** stone. Their **[heaviness | lightness]** surprised the doctor.

Getting exam results.

The girls crowded around the notice-board to see the exam results for chemistry and biology. Only a few of the girls scored more than **[ninety | five]** percent. Their **[joy | grief]** was obvious to everyone.
Were the results for a physics exam?

In the bar.

One of the barstaff liked to have lewd jokes with the female customers when they ordered a drink. Only a few of the women **[enjoyed his jokes | returned to the bar]**. Their **[pleasure | displeasure]** was obvious to everyone.

At the graduation ceremony.

The University's graduation ceremony was always an extremely expensive affair. Only a few of the students could afford to hire **[graduation robes at seven pounds a day | a Rolls-Royce at one hundred pounds a day]**. Their **[wealth | poverty]** shocked the Principal.

Going to a party.

The girls, who were all under fourteen, asked their parents if they could go to a birthday party at a classmate's house. Only a few of the parents allowed their daughter to **[take alcohol with them | go to the party]**. Their **[strictness | laxness]** reflected their own up-bringing.
Were the girls under fourteen?

In a restaurant.

The office party went to an up-market Italian restaurant for their Christmas dinner. Only a few of the office staff **[offered to buy the waiter a drink | ate with their mouths closed]**.
Their **[courtesy | rudeness]** was a surprise to the waiter.
Did the office party go to an Italian restaurant?

At a boy-scout camp.

It was dark before the boy-scouts finished putting up the tents, collected wood for the camp-fire, and sat around the fire to have a meal. Only a few of the boy-scouts went to bed **[early | that night]**. Their **[wakefulness | sleepiness]** annoyed the scoutmaster.
Did the boys build a camp-fire?

Filler materials for Experiment 7
The following are the filler materials for Experiment 7. Some of the items were materials for another experiment. I have listed only those which appeared in one of the presentation blocks.

In a taxi.

A taxi-driver dropped a man and a woman off at the station. On the way they had run into a traffic jam. The taxi driver and the woman had been arguing but the man had remained calm. The taxi-driver told the woman that he had been arguing with that she wouldn't miss the train.

The artist sent the flowers to the teacher
and the joiner brought the whisky for the workmate.

Although the very long film is
frightening the young child the mother enjoys the plot
and although the rather tall man is
singing the theme tune the chorus annoys the crowd.

The major sent the order back to the headquarters.
The model found the jacket changed in the studio.

Although the popular story was
charming the cheerful cast the director was uncertain.
Although the elegant lady was
knitting the turquoise socks were unfinished.
Were the socks finished ?

The programmer posted the software hugged the librarian.
The grandparent knitted the jumper for the adolescent.

Since the very famous case was
intriguing the clever sergeant noted the facts.
Since the rather worried nurse was
bandaging the bleeding patient the consultant waited in fear.

The actor reserved the ticket took the taxi
and the lawyer despatched the report left the airport.

Because the exceedingly long route was
tiring the injured walker stopped for a break
and because the surprisingly keen boy was
ironing the yellow jacket hung on a chair.

The executive sent the blackmail letter to the chairperson
and the ambassador faxed the crucial visa to the traveller.

As the unhelpful reports were
discouraging the surly conscripts the sergeant was annoyed
and as the annoying lawyers were
photocopying the crucial papers the statement was mislaid.

The disc jockey played the song praised the guitarist
and the pop singer sold the car fired the manager.

While the African safari was
exciting the naive tourist flipped through books
and while the Portuguese musician was
composing the splendid opera played for days.

The barrister whispered the verdict coaxed the defendant.
The producer assigned the programme to the cameraman.

As the deafening thunder was
menacing the desperate yachtsman returned quickly.
As the careful rider was
polishing the embossed buckle the harness glinted brightly.

The banker saved the seat next to the client.
The farmer built the house hated the town.

Although the important document was
convincing the stupid boss the youth was not taken in.
Although the aggressive magistrate was
entering the noisy court did not settle down.

In the street.

A burglar looked over most of the houses in the street. Several of them had lights on so he concentrated on the others. Eventually he picked his target and broke in by the rear window.
Were the houses on the same street?

The waiter handed the glass glimpsed the charming hostess
and the chemist rented the flat feared the shady landlord.

While the recently produced video was
boring the old experts watched with great interest
and while the extremely famous scientist was
choosing the best samples cooled in the compartment.

The woman painted the portrait near the house
and the client repaid the money to the bank.

Since the rather formal complaint was
worrying the apprehensive singer the composer tried to help
and since the very anxious doctor was
assisting the incompetent surgeon the inspector asked to go.
Did the composer try to help ?

The psychologist lent the book to the nurse.
The mathematician bought the wine kissed the guest.

Because the additional subject was
confusing the history student the teacher asked a question.
Because the apprehensive toddler was
visiting the generous uncle made a jelly.
Did the student ask a question ?

The climber tossed the rope yelled to the injured mountaineer.
The builder passed the paint up to the spotty apprentice.

Although the particularly loud noise was
distracting the little groups were very good.
Although the unusually smart boy was
explaining the complex rules the game seemed rather hard.
Did the game seem simple ?

The publisher read the manuscript met the editor
and the consultant passed the stethoscope called the researcher.

While the further new evidence was
dismaying the young defendant was looking nervous
and while the rather bold architect was
surveying the whole property was being levelled.

The baby fed the apple to the dappled pony.
The umpire tossed the jumper stopped the cricket player.

Because the extensive landscapes were
enchanting the fat duke the chauffeur was very cheerful.
Because the expanding grocers were
opening the main shop was rather crowded.

The prince served the meal heard the princess.
The witch brought the frog to the wizard.

Since the American patrol was
threatening the cowardly terrorist fled speedily.
Since the Canadian expert was
directing the wonderful musical the performance went perfectly.

The singer returned the records to the confused composer
and the agent offered the booking to the famous magician.

As the lengthy and pretentious review was
insulting the friendly producer the actor contacted friends
and as the eager but unhappy student was
programming the ancient computer the keyboard emitted sparks.

Appendix 4:

Resullts for the off-line post-test conducted following Experiment 7

Subjects made a forced choice between continuations which described a property of either the refset or compset of the previously quantified NP. The table illustrates the number of refset continuations selected when the quantified sentence matched pragmatic norms, and the number of compset continuations which were selected when the quantified sentence violated pragmatic norms. Those items marked with an asterisk were entered into the re-analysis of the Experiment 7 results.

	Norm-matching	Norm-violating
Titles	**Refset Continuation**	**Compset Continuation**
At the Gym	5	10
* In the maths class	6	10
Working in a supermarket	5	10
* At the funfair	6	10
* At the train station	6	11
* Caught by a policeman	8	7
Buying compact discs	4	12
In the train station	5	11
At the swimming pool	5	11
At the circus	4	9
Applying for a job	omitted from analysis	omitted from analysis
Going to a party	10	5
* Comparing heights	10	8
A health inspection	1	9
In the office	2	9
* Running a marathon	11	6
* At the theatre	9	10
* At the boxing match	9	8
* Getting exam results	10	11
* In a bar	9	8
* At the graduation ceremony	6	10
At a party	5	7
* In a restaurant	8	10
At a boyscout camp	5	10